

# Grammatical gender and linguistic complexity

Volume I: General issues and  
specific studies

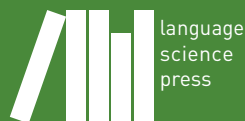
Edited by

Francesca Di Garbo

Bruno Olsson

Bernhard Wälchli

Studies in Diversity Linguistics 26



## Studies in Diversity Linguistics

Editor: Martin Haspelmath

In this series:

1. Handschuh, Corinna. A typology of marked-S languages.
2. Rießler, Michael. Adjective attribution.
3. Klamer, Marian (ed.). The Alor-Pantar languages: History and typology.
4. Berghäll, Liisa. A grammar of Mauwake (Papua New Guinea).
5. Wilbur, Joshua. A grammar of Pite Saami.
6. Dahl, Östen. Grammaticalization in the North: Noun phrase morphosyntax in Scandinavian vernaculars.
7. Schackow, Diana. A grammar of Yakkha.
8. Liljegren, Henrik. A grammar of Palula.
9. Shimelman, Aviva. A grammar of Yauyos Quechua.
10. Rudin, Catherine & Bryan James Gordon (eds.). Advances in the study of Siouan languages and linguistics.
11. Kluge, Angela. A grammar of Papuan Malay.
12. Kieviet, Paulus. A grammar of Rapa Nui.
13. Michaud, Alexis. Tone in Yongning Na: Lexical tones and morphotonology.
14. Enfield, N. J. (ed.). Dependencies in language: On the causal ontology of linguistic systems.
15. Gutman, Ariel. Attributive constructions in North-Eastern Neo-Aramaic.
16. Bisang, Walter & Andrej Malchukov (eds.). Unity and diversity in grammaticalization scenarios.
17. Stenzel, Kristine & Bruna Franchetto (eds.). On this and other worlds: Voices from Amazonia.
18. Paggio, Patrizia and Albert Gatt (eds.). The languages of Malta.
19. Seržant, Ilja A. & Alena Witzlack-Makarevich (eds.). Diachrony of differential argument marking.
20. Hölzl, Andreas. A typology of questions in Northeast Asia and beyond: An ecological perspective.
21. Riesberg, Sonja, Asako Shiohara & Atsuko Utsumi (eds.). Perspectives on information structure in Austronesian languages.
22. Döhler, Christian. A grammar of Komnzo.
23. Yakpo, Kofi. A Grammar of Pichi.
24. Guérin Valérie (ed.). Bridging constructions.
25. Aguilar-Guevara, Ana, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado \*eds.). Definiteness across languages.
26. Di Garbo, Francesca, Bruno Olsson & Bernhard Wälchli (eds.). Grammatical gender and linguistic complexity: Volume I: General issues and specific studies.
27. Di Garbo, Francesca, Bruno Olsson & Bernhard Wälchli (eds.). Grammatical gender and linguistic complexity: Volume II: World-wide comparative studies.

ISSN: 2363-5568

# Grammatical gender and linguistic complexity

Volume I: General issues and  
specific studies

Edited by

Francesca Di Garbo

Bruno Olsson

Bernhard Wälchli


Di Garbo, Francesca, Bruno Olsson & Bernhard Wälchli (eds.). 2019.  
*Grammatical gender and linguistic complexity: Volume I: General issues and specific studies* (Studies in Diversity Linguistics 26). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/223>

© 2019, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/> 

Indexed in EBSCO

ISBN: 978-3-96110-178-8 (Digital)

978-3-96110-179-5 (Hardcover)

ISSN: 2363-5568

DOI:10.5281/zenodo.3446224

Source code available from [www.github.com/langsci/223](https://www.github.com/langsci/223)

Collaborative reading: [paperhive.org/documents/remote?type=langsci&id=223](https://paperhive.org/documents/remote?type=langsci&id=223)

Cover and concept of design: Ulrike Harbort

Typesetting: Bruno Olsson, Sebastian Nordhoff

Proofreading: Ahmet Bilal Özdemir, Andreas Hölzl, Andreea Calude, Brett Reynolds, Calle Börstell, Christian Döhler, George Walkden, Gerald Delahunty, Ivica Jeđud, Jeroen van de Weijer, Kate Bellamy, Lachlan Mackenzie, Laura Arnold, Marc Tang, Ludger Paschen, Martin Haspelmath, Steven Kaye, Tamara Schmidt, Tom Bossuyt, Vadim Kimmelman, Yvonne Treis

Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono

Typesetting software: Xe<sub>La</sub>T<sub>E</sub>X

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

[langsci-press.org](http://langsci-press.org)

Storage and cataloguing done by FU Berlin

Freie Universität  Berlin

# Contents

<b>1</b>	<b>Introduction</b>	
	Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli	<b>1</b>
<b>I</b>	<b>General issues</b>	
<b>2</b>	<b>Canonical, complex, complicated?</b>	
	Jenny Audring	<b>15</b>
<b>3</b>	<b>Gender: esoteric or exoteric?</b>	
	Östen Dahl	<b>53</b>
<b>4</b>	<b>Why is gender so complex? Some typological considerations</b>	
	Johanna Nichols	<b>63</b>
<b>II</b>	<b>Africa</b>	
<b>5</b>	<b>Niger-Congo “noun classes” conflate gender with deriflection</b>	
	Tom Güldemann & Ines Fiedler	<b>95</b>
<b>6</b>	<b>Gender in Uduk</b>	
	Don Killian	<b>147</b>
<b>III</b>	<b>New Guinea</b>	
<b>7</b>	<b>Gender in Walman</b>	
	Matthew S. Dryer	<b>171</b>
<b>8</b>	<b>The gender system of Coastal Marind</b>	
	Bruno Olsson	<b>197</b>

## *Contents*

<b>9</b>	<b>Gender in New Guinea</b>	
	Erik Svärd	225
<b>IV</b>	<b>South Asia</b>	
<b>10</b>	<b>Gender typology and gender (in)stability in Hindu Kush Indo-Aryan languages</b>	
	Henrik Liljegren	279
	<b>Index</b>	329

# Chapter 1

## Introduction

Francesca Di Garbo

Stockholm University

Bruno Olsson

Australian National University

Bernhard Wälchli

Stockholm University

This chapter introduces the two volumes *Grammatical gender and linguistic complexity I: General issues and specific studies* and *Grammatical gender and linguistic complexity II: World-wide comparative studies*.

Grammatical gender is notorious for its complexity. Corbett (1991: 1) characterizes gender as “the most puzzling of the grammatical categories”. One reason is that the traditional definitional properties of gender – noun classes and agreement – are very intricate phenomena that can affect all major areas of language structure. Gender is an interface phenomenon par excellence and tends to form elaborate systems, which is why the question of how systems emerge in language development and change is highly relevant for understanding and modeling the evolution of gender systems. In addition, some of the recent literature on linguistic complexity claims that gender is ‘historical junk’ without any obvious function (Trudgill 2011: 156) and is likely to be lost in situations of increased non-native language acquisition (McWhorter 2001; 2007; Trudgill 1999). Not only are its synchronic functions a matter of debate, but gender also tends to be diachronically opaque due to its high genealogical stability and entrenchment (Nichols 1992: 142; Nichols 2003), making gender a core example of a mature phenomenon (Dahl 2004). However, despite the well-established connection between gender and linguistic complexity, and recent attempts to develop complexity metrics for gender systems (Audring 2014; 2017; Di Garbo 2016) and metrics for addressing



the relationship between gender and classifiers (Passer 2016), there is so far no collection of articles particularly devoted to the relationship between grammatical gender and linguistic complexity.

The two companion volumes introduced here are an attempt to fill this gap. They address the topics of gender and linguistic complexity from a range of different perspectives and within a broadly functional–typological approach to the understanding of the dynamics of language. Specific questions addressed are the following:

- **Measurability of gender complexity:**

What are the dimensions of gender complexity, and what kind of metrics do we need to study the complexity of gender cross-linguistically? Are there complexity trade-offs between gender and other kinds of nominal classification systems? Does gender complexity diminish or increase under the pressure of external factors related to the social ecology of speech communities?

- **Gender complexity and stability:**

How does gender complexity evolve and change over time? To what extent do the gender systems of closely related languages differ in terms of their complexity and in which cases do these differences challenge the idea of gender as a stable feature? How complex are incipient gender systems?

- **Typologically rare gender systems and complexity:**

How do instances of typologically rare gender systems relate to complexity? What tools of analysis are needed to disentangle and describe these complexities?

Discussion around these topics was initiated during a two-day workshop on “Grammatical gender and linguistic complexity” that took place at the Department of Linguistics at Stockholm University, Sweden, November 20–21, 2015. Most chapters included in the two volumes are based on papers first presented and discussed during this workshop. However, some additional authors came on board after the workshop and all contributions went through considerable modifications on their way to being included in the collection of articles. The result consists of 14 chapters (including this introduction) in two volumes, which address the questions listed above, while investigating the many facets of grammatical gender through the prism of linguistic complexity.

The chapters discuss what counts as complex or simple in gender systems, and whether the distribution of gender systems across the world’s languages



relates to the language ecology and social history of speech communities. The contributions demonstrate how the complexity of gender systems can be studied synchronically, both in individual languages and across large cross-linguistic samples, as well as diachronically, by exploring how gender systems change over time.

## Organization of the two volumes

The first volume, *Grammatical gender and linguistic complexity I: General issues and specific studies* (henceforth referred to as Volume I), consists of three chapters on the theoretical foundations of gender complexity, and six chapters on languages and language families of Africa, New Guinea and South Asia. The second volume, *Grammatical gender and linguistic complexity II: World-wide comparative studies* (henceforth referred to as Volume II), consists of three chapters providing diachronic and typological case studies, and a final chapter discussing old and new theoretical and empirical challenges in the study of the dynamics of gender complexity. The rest of this section is a roadmap providing summaries of the following thirteen chapters.

## Volume I: General issues and specific studies

Part I, General issues, in Volume I, starts with **Jenny Audring**'s contribution. Building on previous work in Canonical Typology, Audring proposes that a maximally canonical gender system is one in which formal clarity and featural orthogonality reign, unperturbed by morphological cumulation and cross-category interactions. Canonical gender is also populated by well-behaved targets exhibiting unambiguous agreement, in accordance with the (transparently assigned) gender of their controllers. Alongside this hypothetical clustering of canonical properties, Audring, building on earlier literature, establishes three main dimensions according to which the complexity of a gender system can be gauged: economy (a system with fewer distinctions is less complex than one with many distinctions), transparency (a one-to-one mapping between meaning and form is less complex than a one-to-many mapping) and independence (a system in which all features are independent of each other is less complex than one where they interact). Starting from the postulate that the maximally canonical gender system should also be minimally complex, Audring examines how the canonicity parameters fare against the complexity measures, and finds that the criteria from canonicity and complexity largely converge, with economy being the glaring exception: a canonical gender system is an uneconomical one. The discussion then

turns to the notion of difficulty, here understood as the speed with which children acquire the gender system of their first language. With the premise that a gender system of maximal canonicity and minimal complexity should also be the least difficult to acquire, Audring compares the criteria for canonicity and complexity with factors that are known to facilitate the acquisition of a gender system. The result of this comparison is general convergence between the three dimensions, again except for economy. An otherwise canonical and simple gender system will be easier to acquire if it also features ample redundancy.

Exploring the relationship between language structures and sociohistorical and environmental factors is one of the most debated issues in recent quantitative typological research. In his contribution, **Östen Dahl** asks whether there is a negative correlation between the complexity of grammatical gender and community size in line with the general claim that languages with large populations feature simpler morphology than smaller languages. Gender systems presuppose non-trivial patterns of grammaticalization and complex types of encoding in inflectional morphology. In addition, contact-induced erosion and loss of grammatical gender are well documented in the literature. Yet, Dahl shows that it is very hard to find any clear-cut statistically significant correlation between gender features as documented in the *World atlas of language structures* (WALS) and language size. Similarly, gender features do not clearly correlate with any of the inflectional categories represented in WALS, with the exception of systems of semantic and formal gender assignment, which tend to be found in languages with highly grammaticalized nominal number marking. Dahl argues that in order to better understand the impact that language-external factors may have on the complexity of gender systems, areal and genealogical skewing in the distribution of types of gender systems and the demographic profile of the languages need to be taken into account. Furthermore, he suggests that more elaborate classifications of gender systems than those currently available in typological databases are needed in order to identify those aspects of gender marking that are most likely to adapt to the pressure of language-external factors, as well as a shift in perspective from synchronic to diachronic typologies.

**Johanna Nichols** uses canonicity as a starting point for her discussion of the relative complexity of gender agreement. As in Audring's contribution, exponence of gender is non-canonical inasmuch as it departs from the structuralist ideal of biunique form–function correspondence. Nichols proposes the reasonable hypothesis that gender systems are in fact not complex in themselves. Rather, their complexity is a side-effect of gender arising primarily in languages that have already cultivated considerable complexity elsewhere in their grammars. But empirical testing of this hypothesis suggests that it must be rejected,

because Nichols shows – surprisingly perhaps – that languages with grammatical gender do not display a higher degree of overall morphological complexity than languages without gender. The question is then which diachronic processes cause gender systems to accumulate complexity over time, even when the rest of the morphological system manages to avoid increased complexification. Nichols identifies one clue to this puzzle by comparing gender to participant indexation, and, more specifically, to cases in which such systems display hierarchical patterning (as when a verb form indexes the participant that ranks highest on a hierarchy such as 1, 2 > 3). In Nichols' view, this is an example of a “self-correcting mechanism” that can act as a cap on complexification within indexation systems. Gender systems, on the other hand, do not have recourse to such mechanisms, because markers of gender agreement lack the referential function that participant indexes, such as pronouns, have.

Part II of Volume I focuses on languages of Africa. Gender systems in Niger-Congo languages are among the most studied instances of grammatical gender cross-linguistically. Yet to a large extent this body of research is based on a tradition of analysis which is strongly Bantu-centered and not easily applicable to other language families within and outside Africa. The chapter by **Tom Güldemann** and **Ines Fiedler** seeks to overcome this limitation by proposing a novel toolkit for the analysis of Niger-Congo gender systems. The kit rests upon four notions: agreement class, nominal form class, gender and deriflection, and aims to be universally applicable to the description of any language-specific gender system as well as for the purpose of cross-linguistic comparison. While the notions of nominal form class and agreement class have to do with the concrete morphosyntactic contexts in which nominal and non-nominal gender marking occur, gender and deriflection are more concerned with the abstract, lexical dimension of grammatical gender. By using these analytical tools, Güldemann and Fiedler dismiss the notion of noun class which has been largely used in Niger-Congo studies and which rests on the problematic assumption that there is a systematic one-to-one mapping between nominal form classes and agreement classes. The authors demonstrate the descriptive adequacy of the proposed approach by focusing on data from three genealogically and/or geographically coherent Niger-Congo groups in West Africa: Akan, Guang and Ghana-Togo-Mountain. They show how the new method reveals some important generalizations about Niger-Congo gender systems. For instance, agreement class inventories are always simpler (or at least not more complex) than nominal form class inventories, both in terms of number of distinctions and types of structures. Diachronically, this means that the systems of nominal form classes can be more conservative than those of agreement classes.

The contribution by **Don Killian** discusses the gender system of Uduk, a Koman language of the Ethiopian-(South) Sudanese borderland, with special emphasis on some unusual properties of the agreement and assignment principles operating in the language. Gender agreement in Uduk is primarily realized in a set of clitics that attach to the verb, and which mark the case role and gender of a core argument that immediately follows the verb. The fact that these postverbal clitics only appear when immediately followed by the corresponding argument points to the fundamental role of adjacency in this gender system, a point also illustrated by conjunctions and complementizers, which agree in gender with the following nominal. According to Killian, gender assignment is largely arbitrary, even for the highest segments of the animacy hierarchy, where one could expect to find assignment based on salient features of the referent (such as sex). Furthermore, the irrelevance of the referent for gender assignment extends to pronouns and demonstratives, which invariably trigger agreement according to Class I. Apart from a few formal rules (targeting derived nouns), there seem to be no clear-cut semantic patterns that could bring order to this unwieldy assignment system. Killian proposes that the Uduk gender is non-canonical but relatively simple – features that would easily make this gender system slip under the typologist’s radar.

In the first of three contributions focusing on languages of New Guinea (Part III of Volume I), **Matthew Dryer** presents an overview of gender in Walman, a Torricelli language. Gender agreement in Walman is shown in third person agreement on verbs, where the sets of subject and object affixes distinguish feminine and masculine agreement. Agreement is also found, albeit less systematically, on a subset of nominal modifiers, including some adjectives and demonstratives. Gender assignment is sex-based for humans and large animals, arbitrary for lower animals, whereas almost all inanimates are feminine, with spill-over into the masculine for some natural phenomena (which, like animates, are capable of autonomous force). Dryer presents two analytical puzzles for the description of Walman gender. The first concerns the large group of pluralia tantum nouns, which trigger invariant plural agreement instead of the standard masculine or feminine (singular) agreement. This group of nouns is about twice as large as that of masculine nouns, so if the number of members is taken as decisive for the status of a category, then the pluralia tantum category in Walman is clearly on a par with the two uncontroversial genders. The second puzzle concerns diminutive agreement. The Walman diminutive is not marked on the noun itself (unlike some more familiar derivational diminutives), rather it is realized by dedicated diminutive affixes that replace the usual feminine and masculine gender agreement markers. This makes the diminutive look like an additional gender value,

but Dryer points to the lack of inherently diminutive nouns and the fact that the diminutive sometimes co-occurs with masculine/feminine agreement as good reasons for questioning its status as a gender value. Like other contributions to this book, Dryer's discussion is a good illustration of how interactions between gender and other categories of grammar conspire to make gender systems (as well as the task of analyzing them) more complex.

**Bruno Olsson** shows that the complexity of gender can be addressed from a diachronic point of view by advanced methods of internal reconstruction in the case of a family in which all languages except one are so far poorly documented. The language investigated is Coastal Marind, an Anim language of the Trans-Fly area of South New Guinea. Coastal Marind gender is covert except in a few nouns displaying stem-internal vowel alternation (*anem* 'man [I SG]', *anum* 'woman [II SG]', *anim* 'people [I/II PL]'). Olsson endorses earlier comparative research arguing that vowel alternation within Anim words derives from umlaut triggered by postposed articles inflecting for gender (as they still exist in the perhaps distantly related and areally not too remote Ok languages). By means of statistical analysis, he identifies traces of umlaut for two classes even in non-alternating nouns. The lack of any statistical effect in a third class is explained by class shift of nouns for animals. In Coastal Marind, gender and number are intricately intertwined in an unexpected way. The joint plural of the two animate classes behaves almost identically to gender IV, one of the two inanimate classes (which do not distinguish number). Olsson speculates that gender IV might have originated from pluralia tantum, but since there is no longer a semantic link (no inanimate plural), it is not possible to view gender IV as plural synchronically, despite systematic syncretism with the animate plural throughout a large number of different formal exponents, including stem suppletion. The case of Coastal Marind thus demonstrates that a gender system can become more complex through very specific kinds of interaction with phonology on the one hand and with number on the other.

In the traditional literature on gender, not all continents are equally well represented. New Guinea is a major area that has been notoriously underrepresented so far. **Erik Svärd** investigates gender in New Guinea in an areally restricted variety sample of twenty languages and compares it to gender in Africa and beyond. Unlike Africa, where gender is amply represented in the large language families, the two large families in New Guinea, Austronesian and Trans-New Guinea, mostly lack gender, unlike many small language families and isolates in which gender is attested. As a consequence, gender in New Guinea is diverse and more akin to the global profile of gender in comparison with Africa. Despite the diversity of gender in New Guinea, Svärd is able to identify characteristic properties of

gender in New Guinea. Most languages with gender have a masculine–feminine opposition (where either member can be unmarked), and several gender targets, typically including verbs. Unlike Africa and the Old World in general, formal assignment and overt marking of gender on nouns is rare in New Guinea and, in the few languages having formal assignment, it is usually limited to a subset of the gender classes. However, gender assignment in New Guinea is not typically simple, since many languages have what Svärd calls “opaque assignment”, which does not mean lack of assignment patterns, but rather that exceptions abound. The relevance of size and shape, the existence of multiple noun class systems, and lack of gender in pronouns are further properties characteristic of many languages of New Guinea with gender. Svärd’s comparison of New Guinea and Africa concludes the part on languages in Africa and New Guinea.

In Part IV of Volume I, **Henrik Liljegren** investigates the properties of gender systems and their complexity in 25 of 28 Hindu Kush Indo-Aryan languages. The languages under study are those for which there is enough data in published sources and/or the author’s field data, and are examined against the background of other languages spoken in the area, namely other Indo-Aryan, Nuristani, Iranian, Tibeto-Burman, Turkic and Burushaski. The result is a cross-linguistic survey, which is an intra-genealogical, areal and micro-typological study in one. Despite the close genealogical relationship between the Hindu Kush Indo-Aryan languages, their gender systems are remarkably diverse, ranging from languages with the inherited masculine–feminine distinction pervasively marked on many agreement targets in the southwest (for instance, in Kashmiri) to the Chitral languages Kalasha and Khowar in the northwest, which instead have an innovated copula-based animacy distinction. These two languages also reflect the earliest northward migration of Indo-Aryans in the region. In some languages in the southeast, the sex-based and animacy-based oppositions are combined in concurrent gender systems, as is the case in the Pashai languages and Shumashti, which yield the highest complexity scores among Hindu Kush Indo-Aryan languages. Liljegren shows that the distribution of various kinds of gender systems has both genealogical and areal implications, with different Iranian contact languages in the southeast and southwest yielding a variety of contact effects. Liljegren traces in detail how the entrenchment of gender in this language grouping gradually declines from the southeast to the northwest. Generally in Hindu Kush Indo-Aryan, gender is stable only to the extent that related languages with inherited gender are neighbors. But there are also language-internal factors. The functional load of gender is higher in languages with ergative rather than accusative verbal alignment.

## Volume II: World-wide comparative studies

After having introduced all chapters of Volume I, we now turn to Volume II. To date, the study of gender complexity has largely focused on synchrony. **Francesca Di Garbo** and **Matti Miestamo** demonstrate that diachrony is indispensable for a deeper understanding of the relationship between gender and complexity. They investigate four types of diachronic changes affecting gender systems – reduction, loss, expansion and emergence – in fifteen sets of closely related languages (36 languages in total) from various families and continents. In exploring how the detected types of changes relate to complexity, they find that reduction of gender agreement does not necessarily entail reduction of complexity. Rather complexity can increase both in reducing and emerging gender systems. Across the languages of the sample, there are strong regularities in how different kinds of changes are mapped onto the Agreement Hierarchy. The two opposite poles of the hierarchy, attributive modifiers and personal pronouns, can often be identified as the places of origin for both the decline and rise of gender. Di Garbo and Miestamo argue that two opposite forces, syntactic cohesion and semantic agreement, are at work at the two different poles of the implicational hierarchy. In a similar vein, the two different processes involved in reduction – morphophonological erosion and redistribution of agreement – display different directions of change along the Agreement Hierarchy. Di Garbo and Miestamo consider various cases of language-internal rise of gender and contact-induced gender emergence, and detect striking similarities. The cases under consideration suggest that gender in the process of emergence is non-pervasive and constrained. While gender can disseminate by means of borrowing of lexical items, emergent gender systems in borrowing languages differ in structure from gender systems in donor languages.

Traditional definitions of grammatical gender rely on the notions of noun class, agreement and system. **Bernhard Wälchli** demonstrates that dispensing with these notions and pursuing a radically functional approach to the study of grammatical gender is possible and worthwhile. The chapter is a typological investigation of feminine anaphoric gender grams (as in English *she/her*) in a world-wide convenience sample of 816 languages, based on a corpus of parallel texts (the New Testament). The functional equivalence between the forms extracted from the corpus is ensured by the fact that they cover a single search space across all languages considered. Through this methodology, which is applied to the domain of grammatical gender for the first time, the study finds instances of simple patterns of gender marking in a large number of languages for which no such constructions had been documented before. Three types of simple gender are ex-

tracted from the corpus and analyzed in the paper: non-compositional complex noun phrases, reduced nominal anaphors and general nouns. These instances of simple gender are interpreted as incipient types of gender systems from a grammaticalization perspective. Conversely, cumulation with case in the encoding of grammatical relations is taken as a characteristic feature of complex and mature (i.e. highly grammaticalized) feminine anaphoric gender grams. After discussing the differences between simple and mature gender, the chapter concludes by proposing a functional network for the grammatical gender domain in which the gram approach is reconciled with more traditional approaches based on the notions of noun classes, agreement and system.

While languages can have both gender and classifier systems, the co-occurrence of the two is rare. This suggests that these two different types of nominal classification systems may actually be in complementary distribution with one another. **Kaius Sinnemäki** validates this claim statistically by investigating the distribution of gender and numeral classifier systems in a stratified sample of 360 languages. Complexity is operationalized as the overt coding of a given pattern in a given language and thus, in this case, as the presence of gender and/or numeral classifiers. The study's main hypothesis is that there is an inverse relationship between presence of gender and presence of numeral classifiers. The hypothesis is tested using generalized mixed effect models, which also control for the impact of genealogical and areal relationships between languages on the distribution of the variables of interest. The results reveal a statistically significant inverse relationship between presence of gender and presence of numeral classifier systems and that in addition the two types of nominal classification systems have a roughly complementary areal distribution. Languages spoken within the Circum-Pacific region are more likely to have numeral classifiers than languages spoken outside this area, whereas the opposite distribution applies to gender. This inverse relationship also exists independently of language family and area and thus confirms the study's main hypothesis. According to Sinnemäki, these results, which should be interpreted as a probabilistic rather than an absolute universal, suggest that there is a functionally motivated complexity trade-off between gender and numeral classifiers, whereby languages tend to avoid developing and maintaining more than one system at a time within the functional domain of nominal classification.

The concluding chapter, by **Bernhard Wälchli** and **Francesca Di Garbo**, presents a wide-ranging enquiry into the diachrony and complexity of gender systems, with an emphasis on gender systems as dynamic entities evolving over time. The authors re-examine a variety of phenomena that will be familiar to students of gender, such as gender and the animacy hierarchy, assignment rules, gender agreement, and cumulative expression with other inflectional categories.



But casting the net wider, the chapter also examines various issues that have received less attention in the literature, and which arguably are crucial for understanding the origin, development and synchronic characteristics of gender systems. These include the introduction of inanimate nouns into sex-based gender classes, opaque assignment and the development from semantic to phonological assignment, nouns – and clauses – as targets of gender agreement, and relationships between controller and target that go beyond co-reference and syntactic dependency. Among the 12 sections of the chapter (all of which can be read independently), we also find an exploratory survey of accumulation of nominal marking in the NP (including markers that fall outside the realm of noun classification, such as *one* in the NP *the red one*), and a proposal for a definition of agreement that is intended to capture the fundamental asymmetry between controller and target (as the sites where gender originates and is realized respectively). These and other sections of the chapter question the solidity of some commonly made distinctions, such as that between agreement features and conditions on agreement, or the binary splits between e.g. semantic and formal assignment systems, or the assumption that the category of gender can always be distinguished from that of number. These emerge in a new guise once the dynamic perspective favored by the authors is adopted.

## Acknowledgments

The two volumes are the result of a collaborative endeavor in which not only the editors and authors of the chapters were involved. We would like to thank in particular Yvonne Agbetsoamedo, Jenny Audring, Lea Brown, Greville Corbett, Östen Dahl, Michael Daniel, Deborah Edwards-Fumey, Sebastian Fedden, Jeff Good, Pernilla Hallonsten Halling, Martin Haspelmath, Robert Hepburn-Gray, Dan Ke, Marcin Kilarski, Matti Miestamo, Manuel Otero, Robert Östling, Frank Seifart, Ruth Singer, Krzysztof Stroiński, Anna Maria Thornton and 9 anonymous reviewers, whose comments have contributed to considerable improvement of all chapters. We would also like to thank the editorial board of the series *Studies in Diversity Linguistics* for having supported the book project from its very beginning. In particular, the words of encouragement of the series' editor, Martin Haspelmath, have been very important from the outset to the final stages of the volumes' production. We would also like to thank the Language Science Press team, and in particular Sebastian Nordhoff and Felix Kopecky, for their eagerness in supporting us in all matters of producing open access books. As mentioned earlier, the book project started with the workshop "Grammatical gender and linguistic complexity" which was held in Stockholm, November 20th–21st,

2015. We thank the Department of Linguistics at Stockholm University for hosting the workshop, as well as the workshop participants and the audience for contributing to a very inspiring and stimulating discussion venue. The workshop was ultimately made possible thanks to the vice-chancellor of Stockholm University, Astrid Söderbergh Widding, who, back in 2015, initiated a funding program for collaborative linguistics research between Stockholm University and the University of Helsinki. This funding scheme has since then sparked numerous collaborative projects between the two universities, under the coordination of its scientific committee, directed by Camilla Bardel. The Stockholm-Helsinki cooperation program has through the years led to a large number of joint publications of which the present two volumes are just one example.

## References

- Audring, Jenny. 2014. Gender as a complex feature. *Language Sciences* 43. 5–17.
- Audring, Jenny. 2017. Calibrating complexity: How complex is a gender system? *Language Sciences* 60. 53–68.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
- Di Garbo, Francesca. 2016. Exploring grammatical complexity crosslinguistically: The case of gender. *Linguistic Discovery* 14(1). 46–85.
- McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2-3). 125–166.
- McWhorter, John. 2007. *Language interrupted: Signs of non-native acquisition in standard language grammars*. New York: Oxford University Press.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Nichols, Johanna. 2003. Diversity and stability in language. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 283–310. Oxford: Blackwell.
- Passer, Matthias Benjamin. 2016. *The typology and diachrony of nominal classification*. Utrecht: University of Amsterdam dissertation.
- Trudgill, Peter. 1999. Language contact and the function of linguistic gender. *Poznań Studies in Contemporary Linguistics* 35. 133–152.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. New York: Oxford University Press.

## **Part I**

# **General issues**



## Chapter 2

# Canonical, complex, complicated?

Jenny Audring

Leiden University

Investigating the complexity of grammatical gender begins with the question: What are the dimensions of variation? This question is addressed by Canonical Typology, which provides us with a cross-linguistic road map of gender systems (Corbett & Fedden 2016). Compass and measuring rod are the principles of canonicity, which organise the theoretical space around a canonical centre and then situate real gender systems in this space. In this chapter I compare and contrast the principles of canonicity with those of complexity, and discuss both of them in relation to difficulty. While canonicity, complexity, and difficulty are related notions, it will be argued that they are not identical: individual phenomena can be complex but canonical, or complex but not difficult. The aim of the chapter is to tease apart issues of methodology, description, and theory in order to arrive at a clearer understanding of the complexity of gender.

**Keywords:** gender, complexity, canonicity, difficulty, learnability, economy, transparency, independence, redundancy.

## 1 The typology of gender

### 1.1 Introduction

Typologies are descriptive spaces shaped by the dimensions of cross-linguistic variation. Once laid out, such spaces can be profiled according to various theoretical aims. In the domain of grammatical gender, the best example of this method is the Canonical Typology approach (e.g. Corbett 2006; Brown et al. 2013; Bond 2019; Corbett & Fedden 2016 for gender). By organising the typological variation in gender systems according to the principles of canonicity, we arrive at a better understanding of the feature, from its most canonical manifestations at the centre to the non-canonical systems at the periphery.<sup>1</sup>

<sup>1</sup>For a collection of interesting outlier systems, see Fedden et al. (2018).



The aim of this paper is to further explore the typological space of grammatical gender by comparing and contrasting canonicity with two other evaluative measures: complexity and difficulty.<sup>2</sup> The three notions appear to intersect: one might expect canonical gender systems to be the least complex, and the least complex systems to be the least difficult to acquire or use. However, there are theoretical reasons to assume that canonicity can imply greater complexity, and empirical reasons to believe that lower complexity does not necessarily mean lower difficulty.

The chapter is organised as follows. I first lay out the theoretical perspective taken in this chapter. This section also serves as an overview of the terminology used. Then I introduce the notion of “profiling”, which means organising a typological space according to certain principles. §2 discusses the principles involved in profiling the typology of gender according to canonicity on the one hand and complexity on the other. In §3, I apply the principles to the typological space and compare the results. §4 widens the discussion to cross-linguistic evidence on difficulty in first language acquisition. §5 concludes the paper.

With regard to the three notions compared – canonicity, complexity, and difficulty – the text has an asymmetric structure: canonicity is taken as the baseline for an assessment of complexity, but difficulty is introduced independently and then linked to the other two notions.

## 1.2 Theoretical perspective and terminology

The theoretical perspective taken in this chapter is in line with Corbett (1991; 2013a,b,c). Grammatical gender systems are understood as systems of agreement classes. This means that we follow Hockett’s famous dictum that “[g]enders are classes of nouns reflected in the behaviour of associated words” (Hockett 1958: 231) and take agreement as a definitional property of gender. Nouns serve as agreement *controllers* that determine the form and feature structure of agreeing *target* words. An example is (1) from Italian, where the definite article and the predicative adjective agree in gender with the feminine noun *pasta*.

- (1) Italian (Anna Thornton, p.c.)  
*la pasta è squisit-a*  
DEF.SG.F pasta(F).SG be.PRS.3SG delicious-SG.F  
‘The pasta is delicious.’

---

<sup>2</sup>The terms “canonicity”, “complexity” and “difficulty” are used as technical terms throughout the paper. §2 briefly outlines the relevant theory.

The syntactic configurations in which we find the agreement controller and its targets are called *domains*. The most local domain for gender agreement is the noun phrase (although, of course, finer subdivisions can be made with regard to hierarchical or linear distance within the noun phrase). Many languages, including Italian, show gender agreement in more than one domain. Larger domains are the clause (with predicative agreement targets such as verbs) and the sentence (with relative pronouns as clause-external but sentence-internal agreement targets), but anaphoric agreement can reach beyond the sentence and even span more than one turn in conversation.

The number of different agreement patterns corresponds to the number of gender *values* distinguished in a language (this is less straightforward when languages have inconsistent or mismatching agreement patterns). Gender values often have names, e.g. *feminine* or *uter*, especially in smaller systems with fewer values and when values line up with particular semantic properties. The values in larger systems are commonly labeled by numbers. Some linguistic traditions, e.g. the Bantuist literature, speak of noun classes rather than genders and distinguish numbered singular and plural classes (see example (3) below).

Nouns usually have a consistent gender value as an inherent lexical property. *Assignment rules* that regulate which noun goes with which gender are easy to identify in a number of languages, but less so in others. Such rules can refer to semantic, phonological, or morphological properties of nouns. Consider, for example, the following rules proposed for German (Köpcke 1982: Chapter 3).<sup>3</sup>

- Semantic assignment rule
  - Nouns denoting lexical categories are neuter (e.g. *das Substantiv* ‘the noun’, *das Verb* ‘the verb’, *das Pronomen* ‘the pronoun’)
- Phonological assignment rule
  - Monosyllabic nouns ending in /f/ are masculine (e.g. *der Mensch* ‘the human’, *der Busch* ‘the bush/shrub’, *der Marsch* ‘the march’)
- Morphological assignment rules
  - Nouns that take the plural suffix *-(e)n* are feminine (e.g. *die Tür* ‘the door’, *die Stirn* ‘the forehead’, *die Flut* ‘the flood’)

---

<sup>3</sup>These rules are not categorical but reflect statistical tendencies; counterexamples can be found for every proposed rule.

Phonological and morphological rules are often subsumed under “formal rules” (Corbett 2013c). In addition, as defended in Audring (2017), it may be useful to distinguish between general rules that account for a large part of the noun vocabulary, and ‘parochial’ rules with a narrower scope.<sup>4</sup> This distinction cross-cuts the semantic/formal split. The German examples above represent parochial rules; they constitute a small part of a large and complex rule system.

Taken together, the number and nature of the assignment rules, the properties of the controllers, the range of values, and the behaviour of the targets in each domain can be used to broadly characterise the gender system of a language and compare it to others.

### 1.3 Profiling

In typologies of grammatical (sub)systems, all instances of cross-linguistic variation can be treated equally by simply cataloguing the available options. Table 1, for example, lists a selection of options for gender systems.

Table 1: Possible properties of gender systems (selection)

Controllers:	Noun, pronoun, ...
Targets:	Adjectives, verbs, pronouns, articles, ...
Domains:	Noun phrase, clause, ...
Values:	2 gender values/ 10 gender values, ...
Assignment rules:	Semantic, phonological, ...

However, it might be useful to profile the typology. For example, typologists might sort the various options according to commonness or rarity. Alternatively, we might want a typology of gender to say that a gender system with nothing but pronominal targets is a non-canonical gender system – hence the persistent disagreement in the linguistic literature on whether or not English has grammatical gender.<sup>5</sup> Such differences can be captured by defining a “canonical” or ideal gender system and then situating real systems according to their relative distance from this baseline. This is the method of Canonical Typology (Corbett 2006; 2012; Brown et al. 2013; Corbett & Fedden 2016); we will discuss it in more detail in §2 and §3.

<sup>4</sup>For an insightful discussion of parochial or “crazy” rules and associated theoretical issues see Enger (2009).

<sup>5</sup>See Wälchli (2019 [in Volume II]) for a different view on pronominal gender.



Profiling – be it in terms of commonality, canonicity, or any other evaluative measure – organises the typological space according to certain principles and thereby enriches the description, allowing for a deeper understanding of the grammatical (sub)system in question. In the present paper, I will compare two profiles for grammatical gender, the canonicity profile and the complexity profile, and relate both to the issue of difficulty. First, however, we need to establish principles that allow us to ask which properties count as canonical or complex, and why.

## 2 Principles

### 2.1 Introduction: Principles

The method I have referred to as “profiling” creates organised typological spaces. Organisation requires principles. In this section, I will review the principles of canonicity as proposed in the literature, and then suggest a number of possible principles for complexity and difficulty (again, guided by the relevant literature).

Since the issues are themselves highly complex, the representation will be uncomfortably sketchy in places. Especially for canonicity, the reader is referred to the original sources for a more extensive motivation of the approach, for discussion, and for further examples.

### 2.2 Principles of canonicity

The main purpose of the canonical approach to typology is to define a linguistic equivalent of the zero on the Kelvin thermometer: an absolute calibration point in the space of possibilities (Fedden & Corbett 2015). Unlike the scale of a thermometer, however, a canonical typology is multi-dimensional. Corbett & Fedden (2016) define the calibration point for grammatical gender and the variational space around it with the help of a number of principles. Since gender is a morphosyntactic feature involving agreement, most of the principles for canonical gender systems follow from those for canonical morphosyntactic features (Corbett 2012) and canonical agreement (Corbett 2006), respectively. Corbett & Fedden (2016) present the clusters of principles separately; in the following they will be represented jointly. In order to allow for easier cross-reference to the source, the original numbering is retained. This necessitates a minor adjustment: Principle I for canonical morphosyntactic features appears as Principle Ia, Prin-

ciple I for canonical agreement as Principle Ib. Moreover, I have added names to the principles for easier reference throughout the text.<sup>6</sup>

According to Corbett and colleagues, the relevant principles for canonicity are the following (after Corbett & Fedden 2016):

*Principle Ia: Clarity*

The feature gender and its values are clearly distinguished by formal means.

*Principle Ib: Redundancy*

Canonical gender agreement is redundant rather than informative.

*Principle II: Simple Syntax*

In a canonical gender system, the use of the feature and its values is determined by simple syntactic rules. Canonical gender agreement is syntactically simple.

*Principle III: Exponence*

In a canonical gender system, the feature and its values are expressed by canonical inflectional morphology.

*Principle IV: Orthogonality*

Canonical gender and canonical parts of speech are fully orthogonal.

*Principle V: Matching Values*

In a canonical system of grammatical gender the contextual values match the inherent values.

*Canonical Gender Principle (CGP)*

In a canonical gender system, each noun has a single gender value.

The principles are operationalised by means of criteria that specify for individual properties or behaviour whether they are more or less canonical. Greatly simplifying the complex and sophisticated account in Corbett & Fedden (2016), the principles and criteria for canonical gender say that gender

- should be expressed by means of affixes

---

<sup>6</sup>All principles in this chapter are capitalised.

- should involve dedicated and unique markers that express gender and nothing else
- should be marked consistently, regularly, and obligatorily
- is not impinged upon by syntax, lexical restrictions, or other grammatical features.

Controller and target should

- have gender and express it overtly
- have matching values (thus rendering the gender information on the target redundant).

Furthermore, there should not be any syntactic complications such as inconsistent controllers or special agreement rules for different parts of speech. In principle, all relevant parts of speech should have access to all gender values. The exception is nouns, which – canonically – should only have a single, fixed gender value.

Anticipating a more detailed discussion in §3, let us look again at Italian to see how the principles play out.<sup>7</sup> Example (1) is repeated as (2a); example (2b) is added for contrast.

(2) Italian

- a. *la pasta è squisit-a*  
 DEF.SG.F pasta(F).SG be.PRS.3SG delicious-SG.F  
 ‘The pasta is delicious.’
- b. *il cibo è squisit-o*  
 DEF.SG.M food(M).SG be.PRS.3SG delicious-SG.M  
 ‘The food is delicious.’

Italian marks gender mostly by suffixes, which are consistent, regular, and obligatory. However, some cumulative exponence occurs: the definite articles fuse stem and gender marker, and all gender markers double as number markers. Both controllers and targets distinguish two values (masculine and feminine); these match across domains. The great majority of nouns have a constant gender

---

<sup>7</sup>See Fedden & Corbett (2017: 3) for a similar assessment.

value, and many nouns show their gender overtly. Gender agreement is redundant in most cases. Hence, the Italian gender system comes fairly close to being canonical.

Generalising, we can state that a canonical gender system is defined by formal clarity, syntactic and morphological simplicity, orthogonality to all other compatible linguistic properties, and consistency in the behaviour of all items involved. Viewed in this way, it is easy to see that canonicity involves similar considerations to complexity. Indeed, Principle II (Simple Syntax) makes explicit reference to simplicity. Turning to complexity next, we ask what principles can be brought to bear in order to identify a particular property or behaviour as more or less complex.

### **2.3 Principles of complexity**

The literature on linguistic complexity is vast, and many sources propose principles of complexity. The following section draws on Audring (2017), a detailed study of the complexity of gender systems; the principles are inspired by earlier work, chiefly Kusters (2003), Miestamo (2008), and Di Garbo (2014; 2016). Here, as in most sources (with the exception of Kusters 2003), discussion will be restricted to absolute or descriptive complexity (Miestamo 2008; Sinnemäki 2011; 2014) in order to keep relative complexity, i.e. difficulty, a separate issue (for which see §4).

The most common principle applied in judging complexity is that less equals less complex. This kind of assessment can be used for properties that can be counted or measured. For example, a language with two gender values is less complex than a language with four. Other countable properties are, for example, the number of distinct forms in a paradigm or the number of allomorphs for a given grammatical formative. Following Kusters (2003), this might be called the Principle of Economy (but see Miestamo 2008; Garbo & Miestamo 2019 [in Volume II] who call it “Principle of Fewer Distinctions”) and be defined as follows:

Principle of Economy: The more distinctions or forms a grammatical feature involves, the more complex the feature.

The Principle of Economy needs to be supplemented by other principles, since not all phenomena lend themselves to quantification. For example, it might be argued that dedicated, unique markers are less complex than polyfunctional markers. This is not a matter of quantity, but a matter of mapping function to form. Polyfunctionality comes in various guises; the most common are markers that are

syncretic across gender values or that simultaneously express another grammatical feature. The examples in (3) from Chichewa (Niger-Congo (Bantoid), Bentley & Kulemeka 2001) illustrate both situations.

(3) Agreement in Chichewa

- |    |                                |                                |
|----|--------------------------------|--------------------------------|
| a. | <i>mwa-muna a-kuyimba</i>      | <i>a-muna a-kuyimba</i>        |
|    | 1-man      1-sing.PRS          | 2-man    2-sing.PRS            |
|    | ‘The man is singing.’          | ‘The men are singing.’         |
| b. | <i>chi-patso chi-kugwa</i>     | <i>zi-patso zi-kugwa</i>       |
|    | 7-fruit    7-fall.PRS          | 8-fruit    8-fall.PRS          |
|    | ‘A piece of fruit is falling.’ | ‘Pieces of fruit are falling.’ |

The nominal and verbal prefixes in (3) express noun class as well as number: 1 and 7 are singular classes, 2 and 8 are plural classes. (3b) shows the expected situation: the markers for class 7 and 8 are distinct. In (3a) the verbal prefix is syncretic for singular and plural and hence polyfunctional (the same marker also returns as the marker of the plural class 14; Mchombo 2004: 6).

In order to capture the intuition that polyfunctional markers are more complex than dedicated markers, we assume a principle that is well-represented in the complexity literature, the Principle of Transparency (again, I follow the terminology of Kusters 2003; Miestamo 2008 and Garbo & Miestamo 2019 [in Volume II] call it “Principle of One-Meaning-One-Form”). This principle states that:

Principle of Transparency: Minimal complexity is characterised by a 1:1 mapping of meaning and form.

The examples in (3) violate this principle by showing forms with more than one function (cumulative expression of noun class and number in (3a) and (3b), syncretic markers for class 1 and 2 in (3a)). It should be noted that otherwise the Chichewa examples are remarkably transparent: they involve clearly separable prefixes which are even alliterative between controller and target in class 7, 8 and 2.<sup>8</sup>

Certain cases of polyfunctionality produce complex situations for which it seems justified to posit a separate complexity principle. Following Di Garbo (2014; 2016), I call it the Principle of Independence. This principle states that:

Principle of Independence: In the least complex situation, a grammatical feature is independent of other grammatical features or other linguistic properties.<sup>9</sup>

<sup>8</sup>Corbett (2006: 15) includes alliterative form as a criterion for canonical agreement.

<sup>9</sup>See also Corbett (2012: 170, 174) for related criteria for canonical features.

Independence is compromised when gender marking is neutralised for a part of the paradigm. Well-known examples are gender neutralisation in the plural and in the local persons. Table 2 illustrates the latter case. Ngala (Siewierska 2013, data from Laycock 1965) distinguishes gender in all three persons of the singular personal pronouns, while in Arabic (Ryding 2005: 298–299) only the second and the third person mark gender. Italian shows gender in the third person only.

Table 2: Gender marking in personal pronouns (singular)

Language	Ngala (Sepik)		Arabic (Afro-Asiatic)		Italian (IE)	
	M	F	M	F	M	F
1st person	<i>wn</i>	<i>ñan</i>	<i>anaa</i>		<i>io</i>	
2nd person	<i>mən</i>	<i>yn</i>	<i>anta</i>	<i>anti</i>	<i>tu</i>	
3rd person	<i>kər</i>	<i>yn</i>	<i>huwa</i>	<i>hiya</i>	<i>lui</i>	<i>lei</i>

In Arabic and Italian we see that gender depends on another property, in this case another grammatical feature. According to the Principle of Independence, this represents increased complexity because it necessitates longer descriptions of the system. The idea is the same as limited orthogonality in canonicity (Principle IV (Orthogonality) for canonical morphosyntactic features, §2.2 above): not all logically possible pairings of cross-cutting properties occur. Limitations to Independence can involve properties such as part of speech, other features such as person, number, definiteness, or case, lexical restrictions such as lack of productivity of morphological markers, or interventions from the side of the speaker for semantic or pragmatic purposes.

In contrast to canonicity, where the principles and criteria should converge on the same outcome, the three principles of complexity – Economy, Transparency and Independence – are autonomous and can lead to different evaluations. Consider again the Arabic and Italian paradigms in Table 2. From the perspective of Economy the paradigms are simpler than the paradigm of Ngala: they contain fewer forms. However, they violate Transparency by requiring a non-1:1 mapping of features and forms, as *anaa*, *io* and *tu* have to map onto both gender values.<sup>10</sup>

<sup>10</sup>Note that we are still dealing with grammatical gender here and not just with the sex of the speaker or the addressee. In Hebrew, which has a system similar to Arabic, addressing an inanimate entity (say, an egg rolling off the table or a misbehaving computer) would require the use of a second-person pronoun in the appropriate grammatical gender value (feminine for the egg, masculine for the computer) (Lior Laks, personal communication).

The Arabic and Italian data also show higher complexity from the perspective of Independence, since gender is not fully orthogonal with person.

The upshot is that we cannot speak of the complexity of gender as a unitary phenomenon. Rather, we can employ the three principles (and potentially others) to evaluate observable properties or behaviour. A profiled typology or “complexity space” of gender does not have a single calibration point of minimal complexity. Violations of any of the principles constitute a more complex situation.

Note that we are only considering languages that have a gender system. Hence, we disregard the fact that having gender in the first place complexifies a language. Nor will we ask about a gender system’s usefulness or functionality. Such issues are addressed elsewhere – see for example Nichols (2019 [this volume]) and Sinemäki (2019 [in Volume II]).

### 3 Canonicity vs. complexity

#### 3.1 Profiling

Profiling the typological space by means of the principles introduced above, we can draw up a comparison for canonicity and complexity. This will be done separately for five parameters: the controller (§3.2), the targets (§3.3), the values (§3.4), the domains (§3.5), and the assignment rules (§3.6). In each section, we will ask what properties are more canonical and what properties are less canonical, building on Corbett (2006; 2012) and Corbett & Fedden (2016).<sup>11</sup> Then we will evaluate the options according to the principles of complexity. For reasons of space, only a selection of properties will be discussed; see Audring (2017) for a fuller account. Please refer back to §2.2 and §2.3 for the principles.

#### 3.2 Controller

As we saw in §2.2, the principles of canonicity lead to certain expectations with regard to properties and behaviour. For canonical controllers in gender systems, these are the following.

##### 3.2.1 Controller: canonicity

A canonical controller is present and expresses gender overtly. This is due to Clarity as well as to Redundancy, since an explicit controller renders the agreement redundant. According to Simple Syntax as well as to the Canonical Gender

---

<sup>11</sup>Corbett & Fedden (2016: 514–517) discuss the properties of values under the heading of “Features”.

Principle, the controller should be consistent in the agreements it takes and have a single, lexically specified gender value.

Systems that deviate from these expectations are less canonical. The question to explore here is whether they are also more complex. Let us consider the properties one by one.

### 3.2.2 Controller: complexity

While an overtly present controller may be expected throughout, absent controllers are cross-linguistically common in pro-drop languages. Consider the Spanish example in (4), where the adjective agrees with an implicit third-person controller.

- (4) Spanish  
*está*            *rot-a*  
be.PRS.3SG break-F.SG  
'It/she is broken.'

In terms of complexity, an absent controller increases Economy because the syntagmatic structure is simpler. By contrast, it constitutes a case of higher complexity from the point of view of Transparency, since there is no form that goes with the controller function. Moreover, a controller that is absent in some cases but present in others is at odds with Independence, since its distribution is influenced by other factors, e.g. pragmatics.

Aside from their presence or absence, controllers differ in whether or not they mark gender overtly. The opposite of overt gender is covert gender; languages with covert gender express the feature only by agreement. An example for a language with overt gender is Turkana (Nilotic, examples 5a); a covert system is found in Dutch (examples 5b). Other languages may show intermediate degrees of overtness.

- (5) Overt vs. covert gender
- |   |                |
|---|----------------|
| a. Overt gender (Turkana, Dimmendaal 1983: 224) |                |
| <i>ε-sikin-a</i>                                | <i>a-ηasep</i> |
| M.SG-breast-SG                                  | F.SG-placenta  |
| 'breast'  | 'placenta'     |



b. Covert gender (Dutch)

<i>vloek</i>	<i>boek</i>
curse(c).SG	book(N).SG
‘curse’	‘book’

The nouns in (5a) show overt gender in the form of class prefixes. The nouns in (5b) do not provide any formal indication of gender. Covert gender is more complex from the point of view of Transparency, since covert gender involves function without form. On the other hand, overt marking involves additional morphological material and an additional locus of marking, so it is more complex from the perspective of Economy. Independence is affected when overt marking is subject to conditions. An example can be found in the Khoisan language Sandawe, where gender marking on the noun is restricted to a number of nouns referring to female persons, which constitutes a lexical condition motivated by semantics (Steeman 2011: 57).

The next property to be considered is the behaviour of the controller with regard to its targets. According to both Transparency and Independence, nouns should be consistent controllers that trigger the same agreement on any target under any circumstance. This captures the insight that hybrid nouns such as Dutch *meisje* ‘girl’, which takes neuter agreement on attributive targets and (mostly) feminine agreement on others, are a complexifying phenomenon in a gender system.

According to the Canonical Gender Principle (henceforth CGP), nouns should have only a single gender value each. Thus, a language like Savosavo (Papuan, Wegener 2012), which allows for manipulation of the gender value for pragmatic purposes, constitutes a non-canonical situation (example 6).

(6) Savosavo (Wegener 2012: 64)

*Ai lo tuvi=na ko tuvi k-aughi ngai-sa*  
 this DET.SG.M house=NOM DET.SG.F house 3SG.F.OBJ-exceed big-VBLZ  
*patu.*  
 BG.IPFV

‘This house (M) is bigger than that house (F).’, lit. ‘This house (M) is big exceeding that house (F).’<sup>12</sup>

In the example, the noun *tuvi* ‘house’ is used first with masculine agreements matching its lexical gender, but later with feminine agreements; this has the effect of emphasising, diminutive-like, the smallness of the house.

<sup>12</sup>VBLZ=verbalizing morpheme, BG=background

Languages like Savosavo, which systematically recategorise nouns for evaluative statements about size or merit (Corbett 2014: 123; Di Garbo 2014: 179), are not only less canonical, but also more complex. They violate Transparency by a 1:2 mapping of nouns and genders as well as compromising Independence, as the recategorisation involves semantic or pragmatic factors.

Table 3 collates the controller properties and their evaluation in terms of canonicity and complexity. A tick indicates alignment between maximal canonicity and minimal complexity. A cross indicates canonicity but increased complexity. A dash means that a principle is not relevant. In Table 3 we see that maximal canonicity lines up fairly well with minimal complexity. An exception is Economy disagreeing with Clarity and Redundancy: more formal evidence makes for a clearer and hence more canonical gender system, but at the cost of parsimony.

Table 3: Canonicity and complexity of the controller

The controller...	Economy	Transparency	Independence
...is present (Clarity, Redundancy)	✗	✓	✓
...has overt expression of gender (Clarity, Redundancy)	✗	✓	-/✓
...is consistent in the agreements it takes (Simple Syntax, CGP)	-	✓	✓

### 3.3 Targets

The list of target properties figuring in the canonicity profiling is extensive. In the following I will restrict the discussion to a number of central properties.

#### 3.3.1 Targets: canonicity

Canonically, the gender value of the target is redundant and depends on the gender value of the noun. This is a consequence of the Principle of Redundancy, but it also touches on Orthogonality, as each target should have access to all gender values in the language. Virtually all principles demand that the target has gender values that match those of the controller; the Principle of Matching Values makes this explicit. According to Exponence, gender should be expressed by bound morphology. Moreover, the markers should be uniquely distinguishable across other logically compatible features and their values (Clarity).

### 3.3.2 Targets: complexity

The informativity or redundancy of the gender value on the target can be illustrated with the help of example (7).

(7) French (Françoise Kably, p.c.)

- a. *elle/il*            *est*            *idiot-e/idiot*  
3SG.F/3SG.M be.PRS.3SG stupid-SG.F/stupid.SG.M  
‘She/he is stupid.’
- b. *tu*    *es*            *idiot-e/idiot*  
2SG be.PRS.2SG stupid-SG.F/stupid.SG.M  
‘You are stupid.’

In (7a) the gender agreement on the adjective is redundant given the gender of the pronominal controller. In (7b), by contrast, the second person pronoun does not distinguish gender, so the gender value on the adjective is informative. How does the difference play out in complexity? Obviously, redundancy is a violation of Economy: it is uneconomical to express the same information twice. From the point of view of Transparency, two views are possible. In one sense, redundancy always violates Transparency since the same feature is marked more than once. In this view, the agreement targets formally realise the gender of the noun. However, it might be argued that the agreement targets themselves have gender as a contextual feature (in the sense of Booij 1996), and whatever item has a feature should mark it. This would bring (7a) in line with Transparency after all. Paradigmatically, the evaluation depends on whether one assumes that the French 2<sup>nd</sup> person pronoun is syncretic for the two gender values or does not have gender at all. The first scenario constitutes a disruption of Transparency – a single form with two functions – but the second does not, as the absence of a distinct form would correlate with the absence of a feature. Finally, Independence attributes greater complexity to (7b) than to (7a) since the gender values F and M on the adjective in (7b) have to be inferred from elsewhere, e.g. from the sex of the addressee.

That targets should depend on the controller and match its values syntagmatically follows from the asymmetry of agreement. Note that this is not counted as a violation of Independence, since it is definitional for the controller-target relation. However, any additional dependency or influencing factor constitutes higher complexity in terms of Independence. Two such scenarios deserve discussion. The first is a target having ‘its own opinion’ about value choice and taking

on a different gender value than the controller's. A case in point is semantic agreement, for which Dutch provides examples.

(8) Semantic agreement (Dutch)

*dat meisje dat uh die daar achter het stuur*  
 DEM.SG.N girl(N)SG REL.SG.N eh REL.SG.C there behind DEF.SG.N wheel(N)  
*zat*  
 sit.PST.3SG  
 'that girl who sat behind the wheel'  
 (Corpus Gesproken Nederlands © Nederlandse Taalunie 2014)

In (8) the agreements that go with the neuter noun *meisje* 'girl' have two different values: the demonstrative determiner is neuter, while the speaker first chooses a neuter relative pronoun, then hesitates and picks a common gender form.

Semantic agreement is pervasive in Dutch relative pronouns, personal pronouns, and possessive pronouns (Audring 2006; 2009); the relative likelihood is in line with the Agreement Hierarchy (Corbett 1979). This behaviour makes the system more complex because it involves semantics in a place where only syntax should matter; this is the Principle of Independence. Note that Economy is not affected, since there are no additional markers involved (at least not syntagmatically; for the paradigmatic situation see next paragraph). Neither does semantic agreement – strictly speaking – affect Transparency, as both form and feature value change.

The second deviation from matching values arises when certain targets are paradigmatically unable to match the controller. This happens when the target distinguishes other values than the controller. Again, Dutch can serve as an example for this deviation from the canonical situation.

Most agreement targets in Dutch distinguish two genders, referred to as common (c) and neuter (n) (Table 4). Two targets diverge from this pattern. The personal pronouns and the possessive pronouns show an additional distinction between masculine and feminine that is not available to the other targets nor, arguably, for the nouns.<sup>13</sup> Note that gender agreement is restricted to the singular, so only singular forms are given.

<sup>13</sup>Here we see an example where the agreement class approach mentioned in §1.2 runs into analytical difficulties, as gender affiliation is a function of target behaviour, but the targets do not behave uniformly.

<sup>14</sup> The common gender adjective has the suffix *-e* and the neuter adjective is a bare stem. This formal distinction is restricted to indefinite contexts.

Table 4: Gender agreement in Dutch

Target/ Gender	DEF	DEM	ADJ	REL	PRO	POSS
C/M	<i>de</i>	<i>deze/die</i>	<i>-e</i>	<i>die</i>	<i>hij</i>	<i>zijn</i>
C/F					<i>zij</i>	<i>haar</i>
N	<i>het</i>	<i>dit/dat</i>	$\emptyset^{14}$	<i>dat</i>	<i>het</i>	<i>zijn</i>

The additional masculine/feminine split on the pronouns is a violation of the Principle of Independence, since it depends on the target type what gender values are available. Also, the choice of the pronouns requires external motivation. Again, and rather counterintuitively, Transparency is not affected, as each form viewed in isolation corresponds to a single value (an exception is the syncretism of the masculine and the neuter in the possessives which is not our concern here). From the point of view of Economy, the paradigmatic mismatch involves super-numerary distinctions, hence higher complexity.<sup>15</sup> Note, however, that other languages might show the reverse pattern – individual targets with fewer distinctions – resulting in lower complexity from the point of view of Economy.

The principles of canonicity not only reflect expectations about the gender value of the target, but also about its morphology. In a canonical system, “gender is realised through agreement by canonical inflectional morphology, which is affixal” (Corbett & Fedden 2016: 509). Interestingly, the difference does not affect complexity as we have defined it here. Neither in terms of Economy nor in terms of Transparency or Independence do we see a compelling reason to say that a bound marker is less or more complex than a free marker (this has been pointed out by Leufkens 2014). Hence, such differences do not affect our complexity evaluation.

More relevant for complexity is the final property considered here: the unique distinguishability of gender on the target. Here dedicated markers for gender contrast with portmanteau markers that also express other features (we have seen an example in (3)). The Principle of Transparency decrees that a unique marker constitutes the least complex situation. This is in contrast with Economy, since dedicated markers make for more distinct forms. Transparency, in turn, agrees

<sup>15</sup>One may argue that the reduced paradigm of the attributive targets results in lower complexity from the point of Economy. However, there is little reason to assume that Dutch nouns still distinguish three genders – speakers are no longer able to systematically distinguish masculine from feminine nouns – and the pronouns (including, surprisingly, the neuter) mostly reflect semantic rather than syntactic properties (Audring 2006; 2009). Therefore, it makes sense to say that the pronouns show more gender distinctions than the nouns, a case of increased complexity.

Table 5: Canonicity and complexity of the target

The target...	Economy	Transparency	Independence
...has a gender value that is redundant rather than informative (Redundancy)	✗	✓	✓
...depends for its gender value on the gender value of the noun (Redundancy)	–	–	✓
...has gender values that match those of the controller and of other targets (Redundancy, Simple Syntax, Matching Values, CGP)	–/✓/✗	–	✓
...has bound expression of agreement (Exponence)	–	–	–
...has a gender value that can be uniquely distinguished across other logically compatible features and their values (Clarity)	✗	✓	✓

with canonicity in its preference for unique markers. Moreover, computing the form of a polyfunctional marker involves other features, which violates Independence.

Concluding this brief survey of target properties, we see that complexity agrees with canonicity for many properties (Table 5; again, a tick indicates alignment between maximal canonicity and minimal complexity, cross indicates canonicity but increased complexity, dash means that a principle is not relevant). Other properties leave complexity untouched. Disagreement is found in two cases: redundancy and non-syncretic markers are more complex in terms of Economy. The alignment between matching values and Economy depends on the individual language situation. Note again that the ‘inbuilt’ dependency of the target on the controller is not counted as a violation of Independence.

### 3.4 Values

The values of a feature are inextricably linked to the items that carry them: the controller and the targets. Therefore, most value-related properties have already been touched on in §3.2 and §3.3, and this section can be brief.

### 3.4.1 Values: canonicity

Canonically, values have at least the two following properties. First, for any given controller and its targets, gender values do not vary. This is in line with Redundancy, Simple Syntax, Matching Values, and the Canonical Gender Principle, which say that target values should mirror controller values, and that controllers have gender as a lexical property. Invariance includes independence of other features and their values, as decreed by Clarity and Orthogonality. Second, gender values should form a closed class. This is due to Orthogonality: in a fully orthogonal system of lexical items and grammatical features, only the lexical items constitute an open class (Corbett & Fedden 2016: 502–503). Again, we ask if the canonical situation is also the least complex.

### 3.4.2 Values: complexity

Gender values show variation when they are open to choice or change under the influence of other factors. We saw variable controller gender values in §3.2, example (6), and variable target gender values in §3.3, example (8). A more complex situation is found in Romanian, where gender values appear to vary between singular and plural, as the neuter gender agreements resemble the masculine in the singular and the feminine in the plural (see Corbett 1991: 150–152 for an account in which the situation is interpreted not as a case of variation, but as a system with non-unique markers for the neuter gender).

In all cases we see a violation of the Principle of Independence. Independence supports invariant gender values, as a minimally complex gender system is self-contained and does not require reference to other morphosyntactic features such as number, or to non-syntactic factors such as semantics or pragmatics. Therefore, any variation or choice makes the system more complex.

The second property can be interpreted as concerning the number of gender values in a language. The higher this number (i.e. the closer to an open set), the greater the range of potential combinations of nouns and gender values, which makes it harder to establish orthogonality (Corbett & Fedden 2016: 502–503).<sup>16</sup> In terms of complexity, fewer gender values also mean lower complexity, though for different reasons: Economy says that the simplest system has the fewest values.

Summarising, we see that the properties of the values affect complexity to a limited degree: the first affects Independence, the second Economy; the other

---

<sup>16</sup>In the earlier literature, the number of values was used as a criterion for distinguishing gender from classifier systems, with the expectation that gender values should form a “smallish” set (Dixon 1982; Aikhenvald 2000: 6).

principles are not affected (Table 6). For both properties, however, maximal canonicity coincides with minimal complexity.

Table 6: Canonicity and complexity of the values

The values...	Economy	Transparency	Independence
...do not vary for any given controller and its targets (Clarity, Redundancy, Simple Syntax, Orthogonality, Matching Values, Canonical Gender Principle)	–	–	✓
...form a closed class (Orthogonality)	✓	–	–

### 3.5 Domains

Moving on to domains – the syntactic configurations in which agreement occurs – we can identify three criteria that contribute to higher canonicity and that can be evaluated for complexity.

#### 3.5.1 Domains: canonicity

For domains we can state that the most canonical domain of agreement is the local domain (i.e. within the phrase containing the controller; Corbett 2006: 21). This is due to Simple Syntax. Indeed, the greater the syntactic distance between controller and target, the more linguistic theories are inclined to exclude the relation from agreement (e.g. by speaking of “cross-reference” instead; for discussion see Barlow 1991 and Barlow 1992: 134–152, Corbett 1991, 2001 and 2006, and Siewierska 1999). Moreover, Clarity increases when there are multiple domains, as more domains provide better analytical evidence for the existence of an agreement system. Multiple domains are also favoured by Orthogonality, as orthogonality between words and features increases with more agreement targets and hence more domains.

Corbett & Fedden give a third criterion for canonical gender: “In a canonical gender system the gender of a noun is constant across all domains in which a given language shows agreement” (Corbett & Fedden 2016: 517). As this ties in with the lexically specified, single gender value of the controller, the matching gender values of controller and target, and the invariance of all targets for any



given controller, all of which were covered in the previous sections, we will not discuss this criterion further.

### 3.5.2 Domains: complexity

When we compare canonicity and complexity (Table 7), the question arises whether gender agreement within the noun phrase should also count as less complex. Interestingly, within the realm of descriptive complexity that does not consider potential issues of (processing) difficulty, none of the three complexity principles favours one option over the other. Local agreement is neither more economical, nor more transparent or less dependent than agreement elsewhere.

The second domain-related property concerns the number of domains. In a canonical world, agreement involves not one domain but several. However, neither Transparency nor Independence penalises single domains, and with respect to Economy, each additional domain makes the system larger and therefore more complex. Here we see a clear case where canonicity and complexity disagree.

Table 7: Canonicity and complexity of domains

The domain...	Economy	Transparency	Independence
...is local (i.e. within the phrase containing the controller) (Simple Syntax)	–	–	–
...is one of multiple domains (Clarity, Orthogonality)	✗	–	–

## 3.6 Assignment

Gender assignment rules regulate which gender value is associated with any given noun. Canonicity has little to say about this issue.

### 3.6.1 Assignment: canonicity

Corbett & Fedden list a single assignment-related criterion for canonical gender, which feeds the Canonical Gender Principle: “In a canonical gender assignment system, the gender of a noun can be read unambiguously off its lexical entry” (2016: 520). The authors conclude that assignment based on semantics is the most canonical situation (see Audring 2017: 65, footnote 22, for an argument against

this position). Gender assignment based on formal properties is considered less canonical.

### 3.6.2 Assignment: complexity

Complexity also favours semantic assignment rules, but for different reasons. The argument goes by several steps. In §1.2 we introduced a distinction between general rules and parochial rules. While this distinction is primarily about scope, it also relates to the number of rules that are needed to account for the gender of every noun in the language: general rules cover a large portion of the noun vocabulary, so the system can operate with only a few such rules, whereas parochial rules take care of a smaller subset of the nouns, requiring more rules overall.

Another factor that is relevant for complexity is the variety of rule types. Does a language employ only semantic rules or also formal rules, and if the latter, are these phonological, morphological, or both?

Complexity is minimal if rules are large in scope (necessitating only a small number of different rules) and of a single type. This is due to Economy: fewer rules and fewer rule types are quantitatively simpler. If we link this to the typological finding that semantic rules can occur without formal rules but not vice versa (Corbett 1991: 64, though see Killian 2015 and Killian 2019 [this volume] on the Koman language Uduk, which arguably uses only formal rules), we end up with the situation that complexity favours semantic rules. This is the same outcome as for canonicity, but for different reasons. Table 8 summarises the overlap.

Table 8: Canonicity and complexity of domains

Assignment rules	Economy	Transparency	Independence
The gender of a noun can be read unambiguously off its lexical entry (CGP); assignment rules are entirely based on semantics	✓	–	–

## 3.7 Summary: canonicity vs. complexity

The comparison of properties of gender systems in terms of canonicity vs. complexity is summarised in Table 9. A number of observations can be made. First, there are various properties that are relevant to canonicity but not to complexity, or only to a single complexity principle; these are indicated by dashes. If

Table 9: Canonicity vs. complexity, summary

	Property	Economy	Transparency	Independence
Controller...	...is present (Clarity, Redundancy)	✗	✓	✓
	...has overt expression of gender (Clarity, Redundancy)	✗	✓	-/✓
	...is consistent in the agreements it takes (Simple Syntax, CGP)	-	✓	✓
Target...	...has a gender value that is redundant rather than informative (Redundancy)	✗	✓	✓
	...depends for its gender value on the gender value of the noun (Redundancy)	-	-	✓
	...has gender values that match those of the controller and of other targets (Redundancy, Simple Syntax, Matching Values, CGP)	-/✓/✗	-	✓
	...has bound expression of agreement (Exponence)	-	-	-
	...has a gender value that can be uniquely distinguished across other logically compatible features and their values (Clarity)	✗	✓	✓
	...does not vary for any given controller and its targets (Clarity, Redundancy, Simple Syntax, Orthogonality, Matching Values, Canonical Gender Principle)	-	-	✓
Values...	...do not vary for any given controller and its targets (Clarity, Redundancy, Simple Syntax, Orthogonality, Matching Values, Canonical Gender Principle)	-	-	✓
	...form a closed class (Orthogonality)	✓	-	-
Domain...	...is local (i.e. within the phrase containing the controller) (Simple Syntax)	-	-	-
	...is one of multiple domains (Clarity, Orthogonality)	✗	-	-
Assignment	The gender of a noun can be read unambiguously off its lexical entry (CGP); assignment rules are entirely based on semantics	✓	-	-

dashes are discarded (i.e. if only ticks and crosses are considered), an interesting pattern emerges. Transparency and Independence always line up with canonicity (again, ticks indicate maximal canonicity and minimal complexity). Economy, by contrast, disagrees with canonicity in the majority of the cases. There are only three properties for which the most canonical option is also maximally simple: mismatching values involving reduced values, fewer gender values, and a purely semantic assignment system. For the latter two, however, we saw that canonicity and Economy arrived at the same preference by different arguments (see §3.4.2 and §3.6.2). Hence the alignment is even weaker than Table 9 suggests.

What we see is that canonical gender systems can be complex, which means that there are areas where complexity is expected of – perhaps even inherent to – grammatical gender. The principles most at odds are Clarity and Redundancy on the side of canonicity and Economy on the side of complexity.

Having completed the comparison of canonicity and complexity, we move on to the third issue under consideration: difficulty. §4.1 introduces difficulty and motivates the evidence selected for this paper. §4.2 identifies and discusses factors that influence difficulty in first language acquisition. §4.3 ties together the results and links them to the previous issues, canonicity and complexity.

## **4 Difficulty**

### **4.1 Introduction: difficulty**

In contrast to descriptive complexity, which is an absolute evaluative measure, difficulty is inherently relative: a particular structure is difficult for somebody in the context of some particular task. The experiencer can be a speaker, a hearer, or a learner, and the task can be, for instance, language processing or acquisition. The following section discusses difficulty in the context of first language acquisition. Adult second language acquisition is excluded because it increases the empirical space by many additional variables, chiefly the first language (Does it have a gender system? Are the systems of L1 and L2 similar?), the learner (age, motivation) and the learning context (amount of exposure, explicit instruction or not). This makes it much harder to isolate the specific factors that accelerate or delay acquisition of gender (though see Kusters 2003 for an account of relative complexity, i.e. difficulty, based on second language acquisition).

There is a wealth of literature available on first language acquisition of gender in a variety of languages. Unfortunately, the languages addressed are mostly Indo-European, with the notable exception of Gagliardi & Lidz (2014) on Tsez,

and a number of studies on Bantu languages (Niger-Congo); see Demuth (2003) for an overview.

Comparison is impeded by the diversity of the studies. Differences range from who is tested (single children, groups of children), when they are tested (the ideal period lies between 2 and 8 years, but most studies cover smaller time spans), how the data is collected (in diary studies, in the lab, naturally or experimentally) to what is tested (mostly production, sometimes comprehension) and on what items (often existing nouns, sometimes nonce nouns). Methodological choices have important theoretical consequences. Comprehension can reveal abilities that are not yet apparent in production (see e.g. van Heugten & Johnson 2010), and performance on different types of item might reflect different types of learning. For example, correct use of gender with existing nouns can reflect item-based learning, while the ability to classify nonce words may indicate the successful discovery of assignment rules.

Also, there are differences in what is considered the point of successful acquisition. Correctness levels may vary between nouns and between genders, but also between agreement targets, whereby early success with targets close to the noun may reflect knowledge associated with individual lexemes or even combinations acquired as holophrases, amalgams, or chunks (MacWhinney 1978: 59–60). Many studies adopt Brown's (1973) method of using 90% correctness as threshold: an error rate of less than 10% means that gender has been successfully acquired.

Such difficulties notwithstanding, the various studies present some indications of the properties of a language that aid or hinder the acquisition of its gender system. These will be discussed next.

## 4.2 Evidence from first language acquisition

We assume that ease of acquisition is reflected in speed of acquisition: simple systems are acquired faster and/or earlier.<sup>17</sup> Gender systems appear to be in place around the age of three in most languages reported in the literature. For the purposes of this section, the most relevant studies are those that compare acquisition in two or more languages and report faster or slower success for individual languages (e.g. Mills 1986; Eichler et al. 2013) or that point out significant delays (e.g. Mulford 1985; Blom et al. 2008).

---

<sup>17</sup>It might be desirable to distinguish fast from early acquisition, since delays can be due to maturational constraints or because one property relies on the mastery of another (and once the first property is mastered, the second is acquired fast; thanks to Bernhard Wälchli for pointing this out). However, the evidence provided by the literature – especially with regard to first language acquisition – is usually on absolute time (early/late) rather than relative time (fast/slow), so the distinction has to be disregarded here.

A review of the relevant literature yields a consensus on four general factors that influence the acquisition of gender. These can be subsumed under the terms

- Frequency
- Perspicuity
- Consistency
- Monofunctionality

Note that these factors are the result of observations rather than theoretical stipulations such as the principles used in canonicity and complexity profiling (§2 and §3). Let us consider each in turn.

#### 4.2.1 Frequency

Frequency reflects the number of times a child is exposed to a particular item or structure. Unsurprisingly, a positive effect of higher frequency is reported in a variety of studies. Particularly for the initial stages, acquisition is described as proceeding in a piecemeal, item-based manner. Correct use of gender morphology may initially be tied to specific lexical items or individual agreement markers which are mastered early because they often (co-)occur in the input (e.g. Mariscal 2009; Szagun et al. 2007; Mills 1986: 115). Conversely, patterns may be delayed because they are represented with insufficient frequency. Rodina (2014), for example, reports that Russian children have difficulties with female person names ending in *-ik* or *-ok* and with nouns such as *doktor* ‘doctor’ when referring to a woman. These nouns contradict morphophonological rules (their form suggests masculine gender) in favour of semantics: adult speakers strongly prefer feminine agreement in accordance with natural gender. While children master the formal rules early, the semantically motivated exceptions are discovered late because such nouns are infrequent in the input.

Frequency can affect entire gender values. A well-known case is the neuter gender in Dutch, which is acquired with an astonishing delay: children still show around 25% errors at age 7 (Blom et al. 2008, see also Keij et al. 2012 and references there). This is due to the much lower frequency of neuter nouns in the language, plus a condition on the neuter form of adjectives that restricts its presence in the input (see footnote 14).

Generalising to gender systems as a whole, we see that frequent marking in general paves the way to early acquisition. Szagun et al. (2007) remark that nouns

co-occur with articles in most contexts in German, which ensures early success in acquisition since articles are important gender cues. Eichler et al. (2013) suggest the same correlation for French. Noun class markers in Bantu appear on a broad range of agreement targets in a variety of domains and are therefore highly frequent. Acquisition studies report that they are in place by age 2;6–3 (Demuth 2003), despite the large number of classes and their low degree of semantic motivatedness. By contrast, mastery of the apparently much simpler English gender system is comparatively slow; gender errors with person names are found beyond age 4 and errors with non-persons beyond age 6 (Mills 1986: 91, 103). The main reason is that there are few cues in the input, since agreement is restricted to pronouns.

Taken together, the evidence suggests that the difficulty of acquiring a gender system is influenced by the frequency with which the child hears the nouns in company of agreeing words. The more agreement targets there are in the language, and the higher their frequency in use, the earlier the system is detected and mastered.

#### 4.2.2 Perspicuity

If the morphological markers are the central cues to acquisition, such cues are expected to work best when they are perspicuous and clear. Formal perspicuity can be a function of phonological weight (including stress) and relative distinctness, but also of the degree to which a gender value is expressed by a typical form. Arias-Trejo & Alva (2013), for example, report that Spanish children are able to use gender agreement as a predictor of form-meaning correspondences in novel nouns from an early age onwards; the authors attribute this to the clear presence of the suffixes *-a* (feminine) and *-o* (masculine) in the input.<sup>18</sup> Similarly, the feminine definite article in Italian is acquired before the masculine because it has fewer allomorphs (Pizzuto & Caselli 1992: 514). For the complex morphological paradigms of Bantu, early and error-free acquisition is reported and explained by the perspicuity of the noun class prefixes (Demuth 2003: 213).

Conversely, perspicuity is impeded by syncretism, especially when reaching across orthogonal features. The German definite article *der*, for example, is syncretic for nominative masculine and genitive feminine. Eichler et al. (2013) mention this factor as an explanation for the slower acquisition of German gender

---

<sup>18</sup>Such explanations are interpretations, and the same facts are sometimes presented as evidence for opposing views. Thus, Mariscal (2009) analyses the difference between Spanish *-a* and *-o* as “subtle” and lists it among the properties that hinder rather than help acquisition (148, 149).

as opposed to French gender, the two systems being otherwise similar in complexity. A similar point is raised for Icelandic (Mulford 1985; Levy 1988) where noun-final *-a* and *-i* can be cues for feminine respectively masculine gender, but both endings occur in various places within the complex inflectional class system, which makes it harder for the child to discover the correlation. Here, clarity overlaps with functionality, a point discussed in §4.2.4 below.

There is interesting, though cursory, evidence that affixes might be more easily detectable than non-affixal phonological gender cues, being more perspicuous as a unit. Studies report that, in particular, diminutive affixes facilitate gender acquisition (e.g. Kempe et al. 2003 for Russian and Cornips & Hulk 2008 for Dutch).

Overall, there is a consensus in the literature that children use formal cues earlier or to better effect than semantic cues. This has been reported for Tsez (Gagliardi & Lidz 2014), French (Karmiloff-Smith 1979), Spanish (Pérez Pereira 1991), German (MacWhinney 1978; Mills 1986), and Russian (Rodina 2014; Rodina & Westergaard 2012). The only dissenting study is Mulford (1985), who finds that Icelandic children master semantic cues earlier (though see Pérez Pereira 1991 for methodological criticism). However, Icelandic may be a language in which neither the semantic nor the formal cues are particularly clear, as Levy (1988) hypothesises.

Perspicuity is not necessarily tied to form. Semantic cues to gender can also vary in semantic perspicuity, i.e. salience. Importantly, what is evident or salient for the adult speaker may not be so for the gender-acquiring child. Studies show that even natural gender, which seems an obvious and straightforward semantic parameter, is not apparent in the use of gender morphology by young children (Szagun et al. 2007 for German; Rodina 2014 for Russian; Mills 1986 for English). A similar argument is brought forward by Plaster & Polinsky (2010) to refute the complex semantics suggested by Dixon (1972) and Lakoff (1987) for the gender system of Dyirbal – the proposed system would be unlearnable, since the semantic parameters would not yet be available to the child.

#### 4.2.3 Consistency

The clearest cues to gender are also the most consistent: an ideal cue has a unique form that consistently represents a particular gender value. This holds for morphological markers as well as entire nouns. Consistency is broken by variation. For example, the female names ending in *-ik* or *-ok* discussed by Rodina (2014) contain an inconsistent cue: the suffixes normally indicate masculine gender. However, such nouns are mastered earlier than the *doktor*-type nouns included in the same study. It might be argued that the former represent a lower degree of



inconsistency, as each individual suffixed noun is either masculine or feminine, whereas the latter show variation for every individual noun.

The basic insight for the acquisition of assignment rules is that categorial rules are the easiest to acquire (Mills 1986: 114). Stochastic rules involving inconsistent cues are harder to figure out and appear to be learned later. The relevant parameter is sometimes called *reliability* or *validity* (MacWhinney 1978), a prominent term in the Competition Model by MacWhinney et al. (1989). Highly valid cues have high predictive power by being consistently associated with a certain gender value.

Summing up the three factors discussed so far, gender cues work best when they are “sufficiently frequent, adequately valid and easily perceivable” (Wegener 1995: 68 for German, translation mine). Similar statements are made for Spanish (Mariscal 2009; Pérez Pereira 1991) and Italian (Pizzuto & Caselli 1992: 545). For the purposes of the present study a fourth factor, monofunctionality, is worth singling out, though it is not entirely independent of the previous three.

#### 4.2.4 Monofunctionality

Gender markers are dedicated or monofunctional when they express gender and nothing else. However, many languages have gender markers that are polyfunctional and encode two or more properties. Shared functions are usually other features such as number or case, inflectional class, or definiteness. Any kind of polyfunctionality affects both clarity and consistency.

The clearest evidence that gender acquisition is delayed by the parallel acquisition of case is adduced for German. Eichler et al. (2013) observe that German gender is acquired later than French, Italian, or Spanish gender and attribute this to the influence of case. Bewer (2004) reports an early peak in gender correctness followed by a relapse when case starts to emerge. Conversely, Pérez Pereira (1991) notes that Spanish gender agreement markers are more transparent because they do not vary with case.

In her famous study on Icelandic, Mulford (1985) finds that gender is acquired late, with a particular delay in the discovery of formal cues. An explanation is sought in the polyfunctionality of the markers in the highly complex Icelandic inflectional class system, which obscures the correlations between the nominal suffixes and gender.

The impact of polyfunctionality on acquisition is strongest in cases where the child can be suspected of erroneously associating gender markers with other functional properties. Bittner (2002) suggests that German children might initially regard the masculine definite article *der* as a marker of subjecthood or

agentivity. Dutch children appear to start out assuming that the Dutch article *de* is a definiteness marker, delaying the discovery of gender (Keij et al. 2012; Cornips & Hulk 2008).

Generally speaking, the earlier acquisition of formal cues reported in §4.2.2 interestingly suggests that form-form correlations might be easier to acquire than form-function correlations, especially when various functions employ the same morphological markers.

Closing this section of literature review, two sporadic observations might be worth noting. Firstly, a variety of studies indicate early mastery of agreement in local domains, with more persistent errors in the use of distant targets such as pronouns. This suggests a correlation between difficulty and domains. Secondly, and partly contradicting the previous point, Pizzuto & Caselli (1992: 545) report tendentially better results for bound morphology over free markers in Italian, with verbal inflection being acquired before pronouns and articles. However, there is little evidence for or against this pattern in the other literature consulted. Both points, however, are in line with what might be expected from the perspective of canonicity. This brings us to the final section, which ties together the three domains of evaluation.

### 4.3 Summary: canonicity, complexity, difficulty

Returning to the question we set out with, we can now ask how the factors relevant to difficulty line up with those pertaining to canonicity and complexity. Table 10 summarises the alignment of difficulty on the one hand with canonicity and the three types of complexity on the other. As in the previous tables, ticks indicate alignment (minimal difficulty, maximal canonicity, minimal complexity). Divergences (minimal difficulty, lower canonicity, higher complexity) are indicated by crosses. Dashes mean no alignment since a factor for difficulty is irrelevant to canonicity and/or complexity.

Table 10: Difficulty vs. canonicity and complexity, summary

Difficulty	Canonicity	Economy	Transparency	Independence
Frequency	-/✓	-/✗	-/✗	-
Perspicuity	✓	✓/✗	✓	✓
Consistency	✓	-	✓	✓
Monofunctionality	✓	✗	✓	✓

Starting with frequency, we saw that difficulty introduces parameters into the discussion that are of limited relevance to canonicity or complexity: the usage frequency of nouns and agreeing elements matters only to difficulty. Syntagmatic frequency as dependent on the number of targets, by contrast, is relevant to all three evaluative measures, but in contradictory ways: canonicity leads us to expect several targets in various domains (Principle of Redundancy, Principle of Orthogonality), which violates Economy and potentially Transparency and therefore results in a more complex system.<sup>19</sup> For difficulty, more targets mean greater perspicuity, hence facilitation of acquisition.

Perspicuity, in turn, lines up with Transparency, Economy, and Independence in that a perspicuous, i.e. alliterative, form without allomorphic variants makes for the best gender cue in acquisition, as well as the most transparent and the most economical agreement marker needing the least additional specifications. Such markers are also the most canonical. Similarly, perspicuity is greater in the absence of syncretism, as is Transparency. Economy, on the other hand, might be said to favour syncretism. It might also favour markers that are unstressed or phonologically light, in disagreement with perspicuity.

Not shown in Table 10 is difficulty diverging from both canonicity and complexity in the preference for formal cues over semantic cues in the early stages of gender acquisition. This is surprising, as semantic motivations for gender are more canonical and potentially less complex.

The third factor relevant for difficulty, consistency, is clearly in line with canonicity: canonical agreement controllers, targets, and values are expected to show predictable, consistent behaviour. This is also the least complex situation according to Transparency and Independence. The Canonical Gender Principle, according to which each noun should have a single gender value, also describes the situation of least difficulty, as variation slows down acquisition.

Moving on to the fourth difficulty factor, monofunctional markers are the easiest to learn as well as the most transparent and the most independent. They are also the most canonical, as monofunctionality ensures the unique distinguishability of gender across other features. Again, this contradicts Economy, which might be said to favour cumulative markers or reduced paradigms.

A less expected outcome from the point of view of functionality is, again, that

---

<sup>19</sup> As noted in §3.3.2, the decision for Transparency depends on the theoretical perspective. Are agreement markers seen as redundantly realising the feature of the noun? Then agreement is always a violation of Transparency. Or do the agreement targets in fact express their own contextual feature (although the value is dependent on the noun)? In this case agreement is not necessarily non-transparent.

form-form relations might initially be easier to detect in the input than form-function relations, with functions being figured out at a later stage.

Finally, however, attention should be drawn to a pattern that might be expected but is not found: there is no evidence for slower acquisition of systems with higher numbers of gender values. Studies on Bantu noun class acquisition (summarised in Demuth 2003) report that agreement within the NP (demonstratives and possessives) is in place around age 2;4–2;6, followed by class prefixes on the noun (2;6–2;8 in Siswati and Sesotho, even earlier in Zulu), then verb agreement. The entire noun class system is mastered by age 3. This matches the age of successful gender acquisition mentioned for Italian and Spanish (see the summary in Eichler et al. 2013: 556), despite the fact that these languages have two gender values while the cited Bantu languages have around seven.<sup>20</sup> By contrast, the acquisition of English and Dutch, which have far fewer gender values, shows much slower progress. This indicates that the number of classes, which seems such a central and obvious criterion for complexity (i.e. Economy), is in itself not at all relevant for difficulty. Here, canonicity, which ascribes no special status to the number of values, lines up better with difficulty than does complexity.

Summing up, we arrive at an interesting result. Of the three principles for complexity, Independence makes the most accurate predictions for difficulty: cross-cutting features, inter-feature syncretism, and one feature depending on another hinder acquisition, as does any compromise on consistency.

Violations of Transparency, in turn, make the system harder to acquire when there are fewer forms than functions. This holds both for the syntagmatic and the paradigmatic dimension, i.e. for syncretism as well as for cumulative exponence. However, syntagmatic transparency violations that involve overrepresented, i.e. redundantly repeated markers appear to be beneficial: redundancy increases the perspicuity of gender and thereby aids acquisition.

As in the comparison of canonicity and complexity (§3.7), Economy is the odd one out. Economy does not line up with canonicity, and violations of Economy often help rather than hinder learning. The burden of acquiring additional morphology and a greater range of agreement domains is eclipsed by the benefits in perspicuity and frequency. Even for the number of gender values no negative effect is found.

As a consequence, canonicity ends up a better predictor of difficulty than complexity. Economy, which is not a priority in canonicity, is also not a priority in difficulty. In fact, low economy with regard to syntagmatic exponence turns out to be an advantage.

---

<sup>20</sup>The number is an approximation, as the Bantuist tradition counts singular and plural classes separately and includes locative classes, which leaves some room for analytical variation.

## 5 Conclusions

In this chapter I have compared and contrasted three evaluative measures: canonicity, complexity, and difficulty. By profiling the typological space of grammatical gender in terms of canonicity and complexity, individual linguistic properties are identified as being more or less canonical, and/or more or less complex. The general result is one of agreement: maximal canonicity lines up well with low complexity and minimal difficulty. The notable exception is the Principle of Economy, according to which maximal canonicity often means higher complexity.

The comparison is then extended to difficulty in first language acquisition. The result is similar: difficulty, canonicity, and complexity largely agree, with the exception of Economy. Violations of Economy can go hand in hand with maximal canonicity and early acquisition. This means that structures may be complex but canonical and easy to learn. This is due to the central role of Clarity respectively perspicuity: systems that offer rich cues and stand out in the grammar provide the best evidence for the linguist and for the language-acquiring child.

The study demonstrates that assessing the complexity, canonicity, and difficulty of gender systems requires typological understanding as well as explicit principles for evaluation in order to arrive at a motivated and consistent judgment.

## Acknowledgements

I am grateful to Grev Corbett, Ray Jackendoff, Anna Thornton, and to the editors of the volume for invaluable comments and advice. The support of the Dutch national research organisation NWO (Veni grant #275-70-036) is gratefully acknowledged.

## Special abbreviations

The following abbreviations are not found in the Leipzig Glossing Rules:

BG	background
C	common gender
VBZ	verbalizing morpheme

## References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Arias-Trejo, Natalia & Elda Alicia Alva. 2013. Early Spanish grammatical gender bootstrapping: Learning nouns through adjectives. *Developmental Psychology* 49(7). 1308–1314.
- Audring, Jenny. 2006. Pronominal gender in spoken Dutch. *Journal of Germanic Linguistics* 18(2). 85–116.
- Audring, Jenny. 2009. *Reinventing pronoun gender*. Vrije Universiteit, Amsterdam dissertation.
- Audring, Jenny. 2017. Calibrating complexity: How complex is a gender system? *Language Sciences* 60. 53–68.
- Barlow, Michael. 1991. The agreement hierarchy and grammatical theory. In Laurel A. Sutton, Christopher Johnson & Ruth Shields (eds.), *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistic Society*, 30–40. Berkeley: Berkeley Linguistics Society.
- Barlow, Michael. 1992. *A situated theory of agreement*. New York: Garland.
- Bentley, Mayrene & Andrew T. C. Kulemeka. 2001. *Chichewa*. Munich: Lincom Europa.
- Bewer, Franziska. 2004. Der Erwerb des Artikels als Genus-Anzeiger im deutschen Erstspracherwerb. *ZAS Papers in Linguistics* 33. 87–140.
- Bittner, Dagmar. 2002. Semantisches in der pronominalen Flexion des Deutschen. *Zeitschrift für Sprachwissenschaft* 21(2). 196–233.
- Blom, Elma, Daniela Polińska & Fred Weerman. 2008. Articles, adjectives and age of onset: The acquisition of Dutch grammatical gender. *Second Language Research* 24(3). 297–331.
- Bond, Oliver. 2019. Canonical typology. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*, 409–431. Oxford: Oxford University Press.
- Booij, Geert. 1996. Inherent versus contextual inflection and the split morphology hypothesis. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1995*, 1–16. Dordrecht: Kluwer.
- Brown, Dunstan, Marina Chumakina & Greville G. Corbett (eds.). 2013. *Canonical morphology and syntax*. Oxford: Oxford University Press.
- Brown, Roger. 1973. *A first language: The early stages*. Cambridge: Harvard University Press.
- Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15(2). 203–224.

- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2001. Agreement: Terms and boundaries. In W. E. Griffin (ed.), *The Role of Agreement in Natural Language - Proceedings of the 2001 Texas Linguistics Society Conference, Austin, Texas, 2–4 March 2001* (Texas Linguistic Forum 53), 109–122. Texas.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2012. *Features*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2013a. Number of genders. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/30>.
- Corbett, Greville G. 2013b. Sex-based and non-sex-based gender systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/31>.
- Corbett, Greville G. 2013c. 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. <http://wals.info/chapter/32>.
- Corbett, Greville G. 2014. Gender typology. In Greville G. Corbett (ed.), *The expression of gender*, 87–130. Berlin: Mouton De Gruyter.
- Corbett, Greville G. & Sebastian Fedden. 2016. Canonical gender. *Journal of Linguistics* 52(3). 495–531.
- Cornips, Leonie & Aafke Hulk. 2008. Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research* 24(3). 267–295.
- Demuth, Katherine. 2003. The acquisition of Bantu languages. In Derek Nurse & Gérard Philippson (eds.), *The Bantu languages*, 209–222. Surrey: Curzon Press.
- Di Garbo, Francesca. 2014. *Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of Africa*. Stockholm University dissertation.
- Di Garbo, Francesca. 2016. Exploring grammatical complexity crosslinguistically: The case of gender. *Linguistic Discovery* 14(1). 46–85.
- Dimmendaal, Gerrit. 1983. *The Turkana language*. Dordrecht: Foris Publications.
- Dixon, Robert M. W. 1972. *The Dyirbal language of North Queensland*. Cambridge: Cambridge University Press.
- Dixon, Robert M. W. 1982. Noun Classes. In ‘Where Have All the Adjectives Gone?’ and other Essays in Semantics and Syntax, 159–183. Berlin: Mouton de Gruyter.

- Eichler, Nadine, Veronika Jansen & Natascha Müller. 2013. Gender acquisition in bilingual children: French-German, Italian-German, Spanish-German and Italian-French. *International Journal of Bilingualism* 17(5). 550–572.
- Enger, Hans-Olav. 2009. The role of core and non-core semantic rules in gender assignment. *Lingua* 119(9). 1281–1299.
- Fedden, Sebastian, Jenny Audring & Greville G. Corbett (eds.). 2018. *Non-canonical gender systems*. Oxford: Oxford University Press.
- Fedden, Sebastian & Greville G. Corbett. 2015. *New approaches to the typology of gender*. Paper presented at the workshop “Non-canonical gender systems”, 48th Annual Meeting of the Societas Linguistica Europaea (SLE), Leiden University, 2 September 2015.
- Fedden, Sebastian & Greville G. Corbett. 2017. Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa: A journal of general linguistics* 2(1). 1–47. DOI:10.5334/gjgl.177
- Gagliardi, Annie & Jeffrey Lidz. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90(1). 58–89.
- Garbo, Francesca Di & Matti Miestamo. 2019. The evolving complexity of gender agreement systems. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume II: World-wide comparative studies*, 15–60. Berlin: Language Science Press. DOI:??
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.
- Karmiloff-Smith, Annette. 1979. *A functional approach to child language: A study of determiners and reference*. Cambridge: Cambridge University Press.
- Keij, Brigitta, Leonie Cornips, Roeland van Hout, Aafke Hulk & Joanne van Emmerik. 2012. Knowing versus producing: The acquisition of grammatical gender and the definite determiner in Dutch by L1-TD, L1-SLI, and eL2 children. *Linguistic Approaches to Bilingualism* 2(4). 379–403.
- Kempe, Vera, Patricia J. Brooks, Natalija Mironova & Olga Fedorova. 2003. Diminutivization supports gender acquisition in Russian children. *Journal of Child Language* 30(2). 471–485.
- Killian, Don. 2015. *Topics in Uduk phonology and Morphosyntax*. University of Helsinki dissertation.
- Killian, Don. 2019. Gender in Uduk. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 147–168. Berlin: Language Science Press. DOI:??
- Köpcke, Klaus-Michael. 1982. *Untersuchungen zum Genusystem der deutschen Gegenwartssprache*. Tübingen: Niemeyer.



- Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflections*. University of Leiden dissertation. Utrecht: LOT.
- Lakoff, George. 1987. *Women, fire, and dangerous things*. Chicago: University of Chicago Press.
- Laycock, Donald C. 1965. *The Ndu language family* (Linguistic Circle of Canberra Publications, Series C Vol. 1). Canberra: Australian National University.
- Leufkens, Sterre. 2014. *Transparency in language. A typological study*. Utrecht: LOT Dissertations.
- Levy, Yonata. 1988. On the early learning of formal grammatical systems: Evidence from studies of the acquisition of gender and countability. *Journal of Child Language* 15(1). 179–187.
- MacWhinney, Brian. 1978. The acquisition of morphophonology. *Monographs of the Society for Research in Child Development* 43(1/2). 1–123.
- MacWhinney, Brian, Jared Leinbach, Roman Taraban & Janet McDonald. 1989. Language learning: Cues or Rules? *Journal of Memory and Language* 28(3). 255–277.
- Mariscal, Sonia. 2009. Early acquisition of gender agreement in the Spanish noun phrase: Starting small. *Journal of Child Language* 36(1). 143–171.
- Mchombo, Sam A. 2004. *The syntax of Chichewa*. Cambridge: Cambridge University Press.
- 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.
- Mills, Anne E. 1986. *The acquisition of gender: A study of English and German*. Berlin: Springer.
- Mulford, Randa. 1985. Comprehension of Icelandic pronoun gender: Semantic versus formal factors. *Journal of Child Language* 12(2). 443–453.
- Nichols, Johanna. 2019. Why is gender so complex? Some typological considerations. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 63–92. Berlin: Language Science Press. DOI:??
- Pérez Pereira, Miguel. 1991. Semantic versus formal theories of gender acquisition: A criticism of Mulford's study. *Archives de Psychologie* 59. 3–16.
- Pizzuto, Elena & Maria Cristina Caselli. 1992. The acquisition of Italian morphology: Implications for models of language development. *Journal of Child Language* 19(3). 491–557.
- Plaster, Keith & Maria Polinsky. 2010. Features in categorization, or a new look at an old problem. In Anna Kibort & Greville G. Corbett (eds.), *Features: Perspectives on a key notion in linguistics*, 109–142. Oxford: Oxford University Press.

- Rodina, Yulia. 2014. Variation in the input: Child and caregiver in the acquisition of grammatical gender in Russian. *Language Sciences* 43. 116–132.
- Rodina, Yulia & Marit Westergaard. 2012. A cue-based approach to the acquisition of grammatical gender in Russian. *Journal of Child Language* 39(5). 1077–1106.
- Ryding, Karin C. 2005. *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.
- Siewierska, Anna. 1999. From anaphoric pronoun to grammatical agreement marker: Why objects don't make it. *Folia Linguistica* 33(1). 225–252.
- Siewierska, Anna. 2013. Gender distinctions in independent personal pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/44>.
- Sinnemäki, Kaius. 2014. Complexity trade-offs: A case study. In Frederick J. Newmeyer & Laurel B. Preston (eds.), *Measuring grammatical complexity*, 179–201. Oxford: Oxford University Press.
- Sinnemäki, Kaius. 2011. *Language universals and linguistic complexity*. University of Helsinki dissertation.
- Sinnemäki, Kaius. 2019. On the distribution and complexity of gender and numeral classifiers. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume II: World-wide comparative studies*, 133–200. Berlin: Language Science Press. DOI:??
- Steeman, Sander. 2011. *A grammar of Sandawe. A Khoisan language of Tanzania*. Utrecht: LOT Dissertations.
- Szagun, Gisela, Barbara Stumper, Nina Sondag & Melanie Franik. 2007. The acquisition of gender marking by young German-speaking children: Evidence for learning guided by phonological regularities. *Journal of Child Language* 34(3). 445–471.
- van Heugten, Marieke & Elizabeth K. Johnson. 2010. Gender-marked determiners help Dutch learners' word recognition when gender information itself does not. *Journal of Child Language* 38(1). 87–100.
- Wälchli, Bernhard. 2019. The feminine anaphoric gender gram, incipient gender marking, maturity, and extracting anaphoric gender markers from parallel texts. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume II: World-wide comparative studies*, 61–131. Berlin: Language Science Press. DOI:??
- Wegener, Claudia. 2012. *A grammar of Savosavo*. Berlin: Mouton de Gruyter.
- Wegener, Heide. 1995. *Die Nominalflexion des Deutschen: Verstanden als Lerngegenstand*. Tübingen: Niemeyer.

## Chapter 3

# Gender: esoteric or exoteric?

Östen Dahl

Stockholm University

Although grammatical gender would seem to be a paragon example of a mature phenomenon in the sense of Dahl (2004), it turns out to be hard to establish any correlation to ecological parameters that have been claimed to co-vary with other such phenomena, such as community size and degree of contact. Grammatical gender also does not seem to correlate with morphological complexity in general. Our understanding of these relationships is hindered by the areal and genetic skewings in the distribution of gender and the lack of diachronic data. To understand how the ecological factors influence the growth, maintenance, and demise of gender systems and eventually their synchronic distribution, we have to go beyond the patterns that can be found in typological data bases like *WALS*. In particular, we need to know more about the conditions under which gender systems arise and mature.

**Keywords:** grammatical gender, esoteric niche, exoteric niche, language ecology, morphological complexity, mature phenomenon, areal typology, community size, suboptimal transmission, semantic gender assignment, formal gender assignment.

### 1 Introduction: The esoteric-exoteric distinction and morphological complexity

In recent decades, many authors have suggested that there is a connection between grammatical complexity, in particular morphological complexity, and factors external to the language system, such as community size, the degree of contact with other language communities and the extent to which the language is learnt and used by non-native speakers (see e.g. the discussion in Trudgill 1983 and Dahl 2004). It appears obvious that a language with grammatical gender is *ceteris paribus* more complex than one without grammatical gender, but can we



say anything about the relationship between grammatical gender and the “ecology” of the language, that is, the conditions under which it is used, learnt and transmitted to new users?

Over the last ten years, there have also been attempts to study the relationship between grammatical complexity and language ecology by quantitative methods. Thus, Sinnemäki (2009: 138) finds in a cross-linguistic investigation that there is “a statistically relatively strong association between community size and complexity in core argument marking, measured as adherence to versus deviation from the principle of one-meaning—one form”. In another study, Lupyan & Dale (2010) make a distinction between “languages spoken in the esoteric niche”, i.e. languages with comparatively smaller populations, smaller areas, and fewer linguistic neighbours, and those spoken in the “exoteric niche”, i.e. languages with larger populations, larger areas, and more linguistic neighbours. Basing themselves on data from the *World atlas of language structures* (WALS; Dryer & Haspelmath 2013), they list more than a dozen morphological features which they have found are more common in languages spoken in the esoteric niche:

- case markings
- ergative alignment
- grammatical categories marked on the verb
- person marking on adpositions
- noun/verb agreement
- inflectional evidentiality
- affixal negation
- morphological future tense
- remoteness distinctions in the past tense
- alienability/inalienability distinctions
- optative mood marking
- distance distinctions in demonstratives
- morphological marking of pronominal subjects
- separate associative plurals

An earlier work that also should be mentioned here is Perkins (1992), who found a negative correlation between language complexity as manifested in deictic grammatical distinctions and cultural complexity as measured by a variety of factors, including the size of communities.

## 2 Is grammatical gender correlated with esotericness?

In Dahl (2004), I introduced the notion of MATURITY as applied to grammatical phenomena. A grammatical pattern was said to be mature if it has a non-trivial prehistory in any language where it appears. I argued that in situations of “suboptimal transmission” of languages, mature patterns will be transmitted less easily and will tend to be reduced or eliminated. As one of “the most mature phenomena in language”, I pointed to grammatical gender. The kind of gender systems we see in some of the major European languages arguably passed through a number of intermediate stages before becoming what they are today. Gender is also a category that depends on inflectional morphology and is conspicuously absent from languages that lack it, such as creoles and the isolating languages of South East Asia and West Africa. We would therefore expect gender to be among the features that have a negative correlation with language size and a positive correlation with general morphological complexity.

But it turns out to be surprisingly difficult to find any such correlation. Already Perkins (1992: 157) points to gender in pronouns and verb affixes as lacking the clear negative correlation with cultural complexity that he finds with other grammatical features such as deictic distinctions in demonstratives. Similarly, gender is not among the features listed by Lupyan & Dale (2010) as being correlated with their esoteric/exoteric dimension. Gary Lupyan (personal communication) informs me that while no consistent relationship can be found between population and sex-based gender systems in the data from *WALS*, there is a weak positive correlation between non-sex-based gender and population, that is, the opposite to what could be expected from what has been said above.

I have made some calculations of my own on the data in the three *WALS* chapters on gender systems (Corbett 2013a,b,c), using iterated samples of one language from each of 60 families or 100 genera, and computing the mean and median values for Pearson’s *r* correlating those samples to the logarithm of the number of speakers of each language (using figures from the *Ethnologue*). This essentially confirmed the findings of Lupyan and Dale, including the weak positive correlation for non-sex-based gender.<sup>1</sup>

---

<sup>1</sup>In Dahl (2011), I reported a positive correlation (0.142) between number of genders and number of speakers in the *WALS* data. That calculation was done on the whole sample, however, and thus did not take account of possible areal and genetic biases.

It is questionable if any firm conclusion can be drawn from the last finding. Judging from the data in *WALS*, non-sex-based gender systems are relatively uncommon – Corbett (2013c) classifies 28 out of 112 gender systems (in a sample of 257 languages) as belonging to this type, and of these 18 are from one single family (Niger-Congo). The total number of families where languages with non-sex-based gender are found is seven, which in my view makes the number of independent cases too small to draw any conclusions.

Thus, we can conclude that it is not possible to show from the data at hand that the presence of gender – or specific types of gender – is correlated to ecological factors such as population. Rather, the evidence suggests the absence of any correlation in any direction (or possibly a very weak positive one).

### 3 Grammatical gender and morphological complexity

I said above that everything else being equal, a language with grammatical gender is more complex than one without grammatical gender. It does not follow, however, that gender is correlated with other kinds of complexity. In fact, Nichols (2019 [this volume]) argues on the basis of a sample of 146 languages that there is no significant difference between gender languages and genderless languages in (i) overall complexity; (ii) morphological complexity in general; (iii) degree of inflectional synthesis of the verb.

These findings can be seen as being in line with the lack of a correlation between gender and ecological factors in the sense that a connection between those factors and a large number of features involving morphological complexity has been demonstrated. On the other hand, the findings are puzzling since gender – following Corbett (1991: 4) – is by definition realized as agreement, and agreement, or perhaps better indexation, would normally be manifested in inflectional morphology. Accordingly, gender is not found in languages traditionally classified as isolating, as noted above.

Trying to elaborate on Nichols' findings, I looked for a correlation between gender and any specific inflectional category in the *WALS* data, but did not find anything close to significance, not even with nominal categories such as case and number. Given that gender and number often go together in inflectional systems, the last finding is particularly puzzling. However, the situation is different if we look just at the languages that have both “semantic and formal gender assignment” and plural marking. For the 26 languages in this group for which there is also information on plural marking, 25 have a morphological plural and out of these, 23 languages mark plural obligatorily on all nouns. In other words, if a

language has gender with formal assignment, it will also tend to have a highly grammaticalized nominal number system.

## 4 Areal and genetic skewings in the distribution of gender

What is easily seen in the *WALS* material is that there are strong skewings in the geographical distribution of gender. About two thirds of the languages with gender systems in Corbett's sample are from Africa and Eurasia; the percentage of gender languages among the languages from those continents is 59, compared to 30 in the languages from the rest of the world. Particularly striking is the distribution of languages with "semantic and formal gender assignment" (Corbett 2013c), where as many as 53 of 59 are found in Africa, Europe, and south-western and southern Asia. Furthermore, nearly all these languages belong to three large families – Afro-Asiatic, Indo-European, and Niger-Congo, which also happen to contain many languages with high speaker numbers, and the few remaining languages are either Nakh-Daghestanian or Khoisan.

In view of what was just said, it would be desirable to factor out possible areal influence from the calculations. This however meets with the problem that the ecological factors that we would like to correlate with the presence of gender are geographically skewed to the same degree, and, in fact, in a similar way. Thus, while 53 of the languages from Africa and Eurasia in Corbett's sample have more than a million speakers, there is just one such language (Guaraní) representing the rest of the world (Australia, the Pacific and the Americas). A more generous sampling would turn up a few more, but it would hardly change the general picture. Nevertheless, it is of some interest to see what happens if the languages from Africa and Eurasia are removed from the calculations of correlation. The results differ only marginally from the ones obtained from the total sample, however, and again it may be questioned if the sample isn't simply too small.

The general conclusion seems to be that it is hard to correlate gender to anything at all, at least as long as we restrict ourselves to the data in *WALS*. It would clearly be better to have a larger sample, but it is not obvious that it would help in the end, due to the heavy areal skewings we find both in gender systems and in the ecology of languages.

## 5 The diachronic perspective

Another problem is the limitation to synchronic data. One observation is that the clustering of gender languages in western Eurasia and adjacent areas of Africa

actually grows stronger as we go back in time and the area occupied by the involved families shrinks. Levins (2002: 252) argues that the Indo-European distinction between masculine and feminine probably arose under Semitic influence, and Matasović (2012) thinks that Indo-European may have influenced those Caucasian languages that have genders. In any case, we cannot unreservedly treat the gender systems in Indo-European, Semitic and Nakh-Daghestanian as independent developments.

In this context, it is important to remember that the probability that a given language exhibits a grammaticalized pattern will depend at least on two different parameters: the propensity for the pattern to arise and the propensity for it to be eliminated in one way or another. It has been claimed (e.g. in Dahl 2004: 199) that gender systems are very stable. What we can see in Corbett's sample is that the families in the western Old World where gender systems with formal assignment show up are very homogeneous as to the presence of gender. Looking at the languages of western Europe, one gets the impression that gender is among the last categories to go when a language undergoes general morphological simplification; thus, many Romance and Germanic languages have lost their case systems but kept gender, although in a somewhat reduced form. It is somewhat hard to generalize here, however – Armenian is an example of a language which has lost gender but preserved its case system (see e.g. Kulikov 2006). It can also be difficult to decide if a category has really disappeared – there may be remnants such as the s-genitive in the Germanic languages, or there may be a renewal of a system, as in the Indic languages, where new case systems have appeared. There is no doubt, however, that a gender system may take a long time to develop but that once it has arisen, it can continue to exist for a very long time. This is bound to weaken the synchronic connection between the presence of gender and ecological factors such as population size, as a gender system may be preserved even if the external situation of the language changes. Moreover, although it is well known that gender systems tend to break down in situations of suboptimal transmission, as in creolization, we know less about the ecological conditions that favour the rise of gender systems.

## **6 Developing the typology of gender systems**

It is thus likely that we have to go beyond synchronic typology to arrive at a fuller understanding of the relationship between gender systems and ecological factors. Detailed comparisons of developments within one and the same family (along the lines of Garbo & Miestamo 2019 [in Volume II]) may shed light on



the problem. But we may also need a more elaborate typology of gender systems, for instance by taking into account in a more systematic way the domains where they operate, and also sharpen the definitions of the features currently used to classify gender systems. Thus, we saw above that the gender systems that are labelled as having “semantic and formal gender assignment” both had a specific geographical distribution and a high correlation with highly grammaticalized grammatical number. On the other hand, the classification behind this label is not fully understood. Corbett (1991: 62) notes that in languages with formal assignment of gender, the gender of a noun is often “evident from its form”, and calls this “overt gender”, as opposed to “covert gender”. He says that in an ideal overt system would have “a marker for gender on every noun” and mentions Swahili as an example of a system that approaches this ideal. But this raises the question of what is basic – the marker or the gender. In fact, the borderline between marking gender and being the source of it is quite thin. For Bantu languages to have overt gender it is necessary to consider the prefixes as being parts of nouns. But consider now Khasi (Austroasiatic), which is treated as having semantic gender assignment in Corbett (2013c). In Khasi, nouns are obligatorily preceded by a “pronominal marker”. There are four such markers: *u* masculine, *ka* feminine, *i* diminutive and *ki* plural. The same elements show up as obligatory 3<sup>rd</sup> person subject markers. Nagaraja (1985: 7) says that “[a] noun without a pronominal marker is not possible” but still treats combinations of pronominal markers and nouns as two-word phrases, in order to “facilitate the dealing with the structure of the nouns as such”. If this choice had not been made, Khasi would look as having a mini-version of a Bantu noun class system, with “overt gender”. We meet a rather similar problem in trying to draw a distinction between gender marking and inflectional classes, as argued in Dahl (2000), exemplified by Scandinavian definite articles, which are manifested both as independent words and as suffixes on nouns, but which vary according to gender in a uniform way wherever they occur (see Dahl 2000 for a discussion).

If we question the role of morphemes such as Bantu noun prefixes as the source of gender assignment, we may also have to reconsider the view that gender assignment is generally rule-governed. Both Killian (2019 [this volume]) and Svärd (2019 [this volume]) argue for the significance of “opaque” or “arbitrary” gender, a possibility that has been downplayed in recent decades. It may be noted that the rise of opaque gender assignment can be seen as an indication of the maturity of a gender system, since it is likely to appear at a relatively late stage of development.

## 7 Conclusion

Although grammatical gender would seem to be a paragon example of a mature phenomenon in the sense of Dahl (2004), we have seen that it is very hard to establish any correlation to parameters that have been claimed to co-vary with other such phenomena. To understand how the ecological factors influence the growth, maintenance, and demise of gender systems and eventually their synchronic distribution, we have to go beyond the patterns that can be found in typological data bases like *WALS*. In particular, we would need to know more about the conditions under which gender systems arise and mature.

## References

- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2013a. Number of genders. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/30>.
- Corbett, Greville G. 2013b. Sex-based and non-sex-based gender systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/31>.
- Corbett, Greville G. 2013c. 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. <http://wals.info/chapter/32>.
- Dahl, Östen. 2000. Elementary gender distinctions. In Matti Rissanen, Terttu Nevalainen & Mirja Saari (eds.), *Gender in grammar and cognition*. Vol. 2: *Elementary gender distinctions: Manifestations of gender* (Trends in Linguistics. Studies and Monographs 124), 577–593. Berlin: de Gruyter.
- Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
- Dahl, Östen. 2011. Are small languages more or less complex than big ones? *Linguistic Typology* 15(2). 171–175.
- Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/>.

- Garbo, Francesca Di & Matti Miestamo. 2019. The evolving complexity of gender agreement systems. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume II: World-wide comparative studies*, 15–60. Berlin: Language Science Press. DOI:??
- Killian, Don. 2019. Gender in Uduk. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 147–168. Berlin: Language Science Press. DOI:??
- Kulikov, Leonid. 2006. Case systems in a diachronic perspective: A typological sketch. In Leonid Kulikov, Andrej L. Malchukov & Peter de Swart (eds.), *Case, valency and transitivity*, 23–47. Amsterdam: John Benjamins.
- Levins, Saul. 2002. *Semitic and Indo-European*. Vol. 2: *Comparative morphology, syntax and phonetics*. Amsterdam: John Benjamins.
- Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLOS one* 5(1). e8559.
- Matasović, Ranko. 2012. Areal typology of Proto-Indo-European: The case for Caucasian connections. *Transactions of the Philological Society* 110(2). 283–310.
- Nagaraja, Keralapura Shreenivasaiah. 1985. *Khasi, a descriptive analysis*. Poona: Deccan College Post-Graduate & Research Institute.
- Nichols, Johanna. 2019. Why is gender so complex? Some typological considerations. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 63–92. Berlin: Language Science Press. DOI:??
- Perkins, Revere D. 1992. *Deixis, grammar, and culture*. Amsterdam: John Benjamins.
- Sinnemäki, Kaius. 2009. Complexity in core argument marking and population size. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an eng variable*, vol. i, 126–140. Oxford: Oxford University Press.
- Svärd, Erik. 2019. Gender in New Guinea. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 225–276. Berlin: Language Science Press. DOI:??
- Trudgill, Peter. 1983. *On dialect: Social and geographical perspectives*. New York: New York University Press.



## Chapter 4

# Why is gender so complex? Some typological considerations

Johanna Nichols

University of California, Berkeley  
Higher School of Economics, Moscow  
University of Helsinki

A cross-linguistic survey shows that languages with gender can have very high levels of morphological complexity, especially where gender is coexponential with case as in many Indo-European languages. If languages with gender are complex overall, apart from their gender, then gender can be regarded as an epiphenomenon of overall language complexity that tends to arise only as an incidental complication in already complex morphological systems. I test and falsify that hypothesis; apart from the gender paradigms themselves, gender languages are no more complex than others. The same is shown for the other main classificatory categories of nouns, numeral classifiers and possessive classes. Person, the other important indexation category, proves to be less complex, and I propose that the reason for this is that person, but not gender, is referential, allowing hierarchical patterning to emerge as a decomplexifying mechanism.

**Keywords:** gender, case, numeral classifiers, possessive classes, person hierarchy, referential, inflection, canonical complexity, simplification, diachronic stability.

## 1 Introduction

There can be little doubt that gender systems are complex, and in various ways: compare the large number of gender classes in Bantu languages, the intricate and opaque fusion with case, number, and declension class in conservative Indo-European languages, the extensive allomorphy of Tsakhur gender agreement



(Nakh-Daghestanian; examples below), or the semantically unpredictable genders of Spanish or French nouns. Even for Avar (Nakh-Daghestanian), which has a three-gender system with almost no allomorphy of gender markers and complete semantic predictability, there is a random division of verbs into those that take gender agreement and those that do not. The open question about the complexity of gender systems is why? Here I propose an answer based on two factors: one is the inexorable growth of complexity as a maturation phenomenon that can continue indefinitely unless braked by some simplification process (Dahl 2004; Trudgill 2011), and the other is a self-correcting measure that is available to some agreement categories but not to gender, for reasons probably having to do with referentiality.

Two different ways of measuring and comparing complexity will be used here. The first is what I will call *inventory complexity*, which goes by various names (e.g. Dahl 2004: *resources*, Miestamo 2008: *taxonomic complexity*, Garbo & Miestamo 2019 [in Volume II]: *the principle of fewer distinctions*): the number of elements in the inventory or values in a system, for some domain such as the number of phonemes, tones, genders, classifiers, derivation types, basic alignments, or basic word orders, or the degree of verb inflectional synthesis. Inventory complexity figures in Dahl (2004), Shosted (2006), Nichols (2009), Donohue & Nichols (2011), and many other works. It is not a very accurate or satisfactory measure of complexity, not least because it does not measure non-transparency, which is the kind of complexity that has been shown to be shaped by sociolinguistics (Trudgill 2011); but it is straightforward to calculate (though data gathering can be laborious), and appears to correlate reasonably well with other, better measures of complexity. Below I use inventory complexity to compare complexity levels of different languages for the practical reason that there is an existing database of inventory complexity (that of Nichols 2009, subsequently expanded) which counts items across several phonological, morphological, and syntactic subsystems across 200 languages.

The other measure used here is *descriptive complexity* or *Kolmogorov complexity*: the amount of information required to describe a system. This is a better measure and captures well the non-transparency relevant to learnability and prone to be shaped by sociolinguistics, but it is very difficult to measure and compare. Here I follow Nichols (2016; forthcoming) in using canonicity theory (Corbett 2007; 2013; 2015; and others) as an approximate measure of descriptive complexity (though not an exact equivalent; some differences are noted below); see Audring (2017) for a similar approach. Canonicity theory is not primarily a

complexity measure but a theoretical undertaking that aims at improving definitions and technical understanding of linguistic notions. It defines a logical space (for a linguistic concept or structure or system) by determining the central, or ideal, position in that space and attested kinds of departures from that ideal, and measuring non-canonicity as the extent of departure (or number of departures) from the ideal. A central notion in defining the ideal position is the structuralist notion of biuniqueness, or one form, one function; any departure from that ideal is non-canonical. The literature on canonicity offers a good deal of work on morphological paradigms, which makes it a straightforward matter to count the number of non-canonicalities in a paradigm. I use canonicity theory partly because of the availability of this previous work and partly because it is well grounded in morphological theory (and taken seriously by theoreticians) yet applicable on its own without requiring adoption of an entire comprehensive formal framework. I survey this kind of complexity with a different database that samples morphological subsystems as sparingly as possible in order to keep the survey manageable (underway; 80 languages so far).

In what follows I illustrate descriptive complexity with some inflectional paradigms and show how much information grammars need to present (and do present) to adequately describe some of those paradigms (§2); this shows that the presence of gender in a paradigm can make it extremely complex by the inventory metric. But is it the gender morphology itself that is complex? Or is gender rather an epiphenomenon of overall language complexity, a category that tends to arise only as an incidental complication in already complex morphological systems? §3 and §4 raise and falsify the hypothesis that gender – and classification more generally – is embedded primarily in already complex languages, showing that it is gender itself that is complex. §5 compares the complexity levels of person, the other important indexation category. It appears that descriptive complexity easily becomes great in the indexation categories, and that person has recourse to self-correcting, self-simplifying mechanisms that gender lacks. More precisely, person has means of self-correction and self-simplification other than sheer reduction of inventory size or overall loss of the category – apparently unlike gender. This partly accounts for the great diachronic stability of gender systems (Matasović 2014) and in particular the remarkable stability of complexity in gender systems. The reason for the different behavior of gender and person appears to be that person, but not gender, is referential. The concluding section (§6) considers some ramifications of this claim.

## 2 Complexity in gender: Examples and measurement

Gender systems can be complex in themselves and also in the way that they interact with other inflectional categories. This section compares some more and less complex gender systems and proposes a way to quantify their complexity. Examples come from the database of non-canonicity, which samples small but easily comparable inflectional subsystems from a few basic parts of grammar in order to get some view of complexity across the inflectional system: marking of A, S, O, G, T, and possessor roles on nouns; the same forms of inflectional pronouns; singular A and O marking in the most basic past and nonpast synthetic forms of verbs; inflectional classes of affixes for nouns, pronouns, and verbs; and inflectional classes of stems for all three.

The paradigms in Tables 1–2 show the inflection of nouns in four grammatical cases in the singular of Mongolian (which has no gender) and Russian (which has three genders).

Table 1: Mongolian (Khalkha; Svantesson 2003: 163, Janhunen 2012: 297–298, 106–112, 66–68; Janhunen’s transcription). Extension underlined.

	‘book’	‘year’
Nominative	<i>nom</i>	<i>or</i>
Genitive	<i>nom-j̄n</i>	<i>or-<u>n-ṽ</u></i>
Accusative	<i>nom-j̄g</i>	<i>or-j̄g</i>
Dative	<i>nom-d</i>	<i>oro-<u>n-d</u></i>

Table 2: Russian (M = masculine, F = feminine, N = neuter). Extension underlined.

	‘brother’ M.anim.	‘house’ M.inan.	‘book’ F	‘window’ N	‘net’ Fourth	‘time’ Fourth, Extended
Nom.	<i>brat</i>	<i>dom</i>	<i>knig-a</i>	<i>okn-o</i>	<i>set’</i>	<i>vremja</i>
Gen.	<i>brat-a</i>	<i>dom-a</i>	<i>knig-i</i>	<i>okn-a</i>	<i>set-i</i>	<i>vrem-<u>en-i</u></i>
Acc.	<i>brat-a</i>	<i>dom</i>	<i>knig-u</i>	<i>okn-o</i>	<i>set’</i>	<i>vremja</i>
Dat.	<i>brat-u</i>	<i>dom-u</i>	<i>knig-e</i>	<i>okn-u</i>	<i>set-i</i>	<i>vrem-<u>en-i</u></i>



#### 4 *Why is gender so complex? Some typological considerations*

Mongolian has only one declension class in terms of suffixes. There are some differences in suffixes (not shown), all predictable from the phonology of the stem (its final consonant and vowel harmony class). There are two stem classes: simple nouns as in 'book', and one with an *-n-* extension in certain cases, as in 'year'. In Russian matters are more complex. There are four declension classes of suffixes: those of 'brother' and 'house', 'book', 'window', and 'net' and 'time' in Table 2, plus a class of indeclinables not shown.<sup>1</sup> There is a minor class of stems with extensions, illustrated here with the *-en-* extension of 'time'. The animate and inanimate masculine nouns differ in their accusative allomorphs; they are largely predictable from the animacy of the referent. Further subclasses not shown here are mostly phonological and predictable from the final consonant or stress position of the stem. (Plural forms and the other oblique cases, not part of this survey, would add further non-canonicalities.)

In canonicity theory, declension classes are non-canonical because they contribute nothing; the one-form-one-function ideal is violated because a declension class has form but no function. There are two kinds of inflectional classes: those involving stems and those involving the inflectional affixes (Bickel & Nichols 2007: 184). Traditionally recognized inflectional classes may be based on stems, affixes, or both, but I factor these out here. A stem declension class has stem change or extension which is a form without meaning; a declension class of affixes is a set of forms but the set has no meaning. The canonical situation is to have no declension classes, so Mongolian is canonical as to affixes (and nearly so as to stems) but Russian is not. On the other hand, if there are declension classes, then they should all be different, since the point of declension classes is differentiation. Affix classes should have affixes all of which are different from the affixes of other classes; each stem class should have an extension, ablaut, stress shift, or whatever that is unique to it. Here Russian declension is non-canonical because there are a number of syncretisms between classes, e.g. the *-u* dative of masculine and neuter declensions or the *-i* genitive of feminine and fourth declensions. Furthermore, within declension classes case affixes should all be different from each other, with one affix per case. Here Russian declension is non-canonical because

---

<sup>1</sup>For this breakdown of the Russian declension classes see Corbett (1982). The traditional terminology deals only with declension classes of endings and not with stem classes. The first three classes are now, at least in work in English, commonly called masculine, feminine, and neuter for the noun genders prototypically or exclusively associated with their members: masculines are only masculine, feminines mostly feminine, neuters only neuter. There is no standard synchronic term for the class of 'net' and 'time'; I call it the fourth declension. Traditionally, the masculine and neuter classes have been grouped together for historical reasons: both go back to the Indo-European *o*-stem declension. The traditional terms are first declension (masculine and neuter), second (feminine), and third ('net' and 'time').

there are many syncretisms within paradigms, such as genitive and accusative for masculine animates or genitive and dative in ‘net’ and ‘time’ in Table 2. A different departure from the principle of a single affix per case is the allomorphy of the accusative ending in the masculine declension: *-a* for animates but zero for inanimates. This is a split of one category into two forms, sensitive to some additional category.<sup>2</sup> (For the general claims of canonicity theory in this paragraph see Corbett 2007; 2013; 2015.)

Thus, of the forms surveyed here, while Mongolian case inflection has one morphological non-canonicity in the system, Russian has 11: the intra-paradigm syncretisms of masculine animate genitive-accusative, inanimate nominative-accusative, neuter nominative-accusative, fourth declension nominative-accusative and genitive-dative; the *-en-* extension in ‘time’; the allomorphy of suffixes between animate and inanimate masculines; and the inter-paradigm syncretisms of nominative zero suffix (masculine and fourth), genitive *-a* (masculine, neuter), genitive *-i* (feminine, fourth), and dative *-u* (masculine, neuter).<sup>3</sup> Both languages have further non-canonicalities in parts of their noun inflectional paradigms that are not surveyed here.

The common types of non-canonicalities in inflectional paradigms are listed in Table 3. All depart from the ideal of one form, one function.

The complexity measurements for the Mongolian and Russian systems shown above are given in Table 4 and Table 5. They pertain only to singular declension; in Mongolian the plural adds no more non-canonicalities, as in the separative morphology of the language plural and case are marked by different morphemes (and the case suffixes are largely the same as in the singular), while in Russian plurality and case are coexponential, with a single suffix signaling the two categories.

Thus a descriptively and theoretically adequate synchronic grammar of Mongolian needs to display only two paradigms, while for Russian five must be shown.

<sup>2</sup>Whether there is a category of animacy that these case forms signal, mark, etc. or they are sensitive to animacy but do not carry it as a category meaning is a thorny issue that cannot be solved here. I will speak of sensitivity to a category (or indeed a property that is not necessarily an actual category of the language) without taking a stance on the larger issue.

<sup>3</sup>Since the extensions of Mongolian appear in some but not all non-nominative cases, perhaps that distribution should also be counted as a non-canonicity, giving Mongolian a total of two. The non-predictability of the Mongolian extension is greater than for Russian: it appears in some but not all non-nominative cases, while the Russian one can be analyzed as appearing in all non-nominative cases (with that pattern then overlain by the nominative-accusative syncretism, which gives an unextended stem to the accusative as well). It is, incidentally, coincidence that the extension has the same consonant in the two languages and appears in the same cases of the partial paradigms shown in Table 1 and Table 2.

#### 4 *Why is gender so complex? Some typological considerations*

Table 3: Non-canonicalities in inflectional paradigms, and their numbers of forms and functions. 2 (+): two or more. 0\*: perhaps defectivity involves not a zero function but an actual function that is blocked from realization.

	Forms	Functions
Syncretism	1	2 (+)
Zero affixes	0	1
Fused exponence (coexponence) of categories	1	2 (+)
Allomorphy, splits	2 (+)	1
Defectivity (gaps)	0	0*

Table 4: Inventory complexity for Mongolian and Russian singular core grammatical cases

	Declensions	Genders
Mongolian	1	0
Russian	5	3, plus animacy

Table 5: Descriptive complexity for Mongolian and Russian singular core grammatical cases. The phonological information is the description in the phonology of automatic alternations.

Mongolian noun paradigms	Russian noun paradigms
Display 1 paradigm, plus 1 extended	Display 5 paradigms, plus extended (2 extension allomorphs)
Access phonological information	Access phonological information Comment on syncretisms, allomorphy, etc.

Pedagogical grammars will usually display more, and, at least for Russian, automatic phonological and morphophonological alternations involving plain vs. palatalized stem-final consonants trigger orthographic changes and are usually also included in the paradigm display. I will not attempt to measure the amount of information presented in the commentaries, notes, etc. on declension paradigms in the two languages, but at first glance it appears to be no less extensive per declension class for Russian than for Mongolian. In any event the difference of one vs. five paradigms suffices to show that more information is required for describing noun declension in Russian than Mongolian.

Russian declension is more complex than Mongolian declension because late Proto-Slavic fused into single case suffixes what had been a sequence of separate stem-forming suffixes (essentially, extensions) plus what had been a more uniform set of case endings in late Proto-Indo-European. The IE extensions had some correlation with gender, and this has tended to increase over time in the attested daughter languages, spurred in no small part by the fact that gender agreement was signalled in adjectives by shifting back and forth between what were lexical or word-formation categories for nouns: *o*-stem suffixes were used for masculine and neuter agreement, the *a*-stem suffixes for feminine. This means that the fusion of gender into the case-number paradigms, an accident of Proto-Slavic sound changes, received support in the gender agreement paradigms of adjectives. This seems to have stabilized the system despite the non-transparency introduced by adding gender to the mix.

Now consider what makes for complexity in a gender system with no fusion of categories or markers. Table 6 shows the gender class markers for Ingush, a Nakh-Daghestanian language of the central Caucasus. Every noun belongs to a gender (usually covert on the noun) marked by root-initial agreement on some verbs and adjectives. Nouns and pronouns referring to male humans belong to V gender, females to J gender; this is what I will call referent-based gender assignment,<sup>4</sup> where gender is predictable from (in this case) the sex of the referent. In the plural both take B agreement, except that first and second person pronouns take D in the plural.<sup>5</sup> Other nouns are arbitrarily assigned to one or another of B, J, and D gender. Altogether there are eight gender classes consisting of singular-plural pairs, and four gender markers. The gender markers have no allomorphy (other than the split of singular B gender into D and B plurals, for which allomorphy is one possible analysis) and no fusion with other segments or morphemes, and are

---

<sup>4</sup>This is the *referential gender* of Dahl (2000). I use *referential* in a different sense; see note 14 below.

<sup>5</sup>In recent linguistic work on Nakh languages the genders are named for the letter name of their marker.

#### 4 Why is gender so complex? Some typological considerations

Table 6: Ingush gender markers (Nichols 2011: 144)

	Singular	Plural	Examples
1st, 2nd person pronouns	<i>v/j</i>	<i>d</i>	me, you, us
3rd person pronouns (human)	<i>v/j</i>	<i>b</i>	him, her, them
male human nouns	<i>v</i>	<i>b</i>	man, Ahmed
female human nouns	<i>j</i>	<i>b</i>	woman, Easet
some animals, inanimates	<i>b</i>	<i>d</i>	ox, head
some plants, inanimates	<i>b</i>	<i>b</i>	apple, family
inanimates, some animals	<i>j</i>	<i>j</i>	wolf, fence
inanimates, some animals	<i>d</i>	<i>d</i>	dog, house

Table 7: Gender agreement in two Ingush verbs. A dot segments off the gender marker. Verbs shown in the simple present tense. (D gender is the citation form.)

	<i>d.ou</i> z- ‘know’ (kennen)	<i>dwa=chy=d.uoda</i> ‘go down’
V	<i>v.oudz</i>	<i>dwa=chy=v.uoda</i>
J	<i>j.oudz</i>	<i>dwa=chy=j.uoda</i>
B	<i>b.oudz</i>	<i>dwa=chy=b.uoda</i>
D	<i>d.oudz</i>	<i>dwa=chy=d.uoda</i>

thus formally transparent. Semantically, as in nearly all gender systems, gender is transparently predictable (referent-based) for nouns and pronouns referring to humans but arbitrary, i.e. opaque, for others.

Formal simplicity vs. complexity is illustrated by the verb paradigms for Ingush and Tsakhur (another Nakh-Daghestanian language: Daghestanian branch, Lezgian subbranch) in Table 7 and Table 8. In Ingush the system is quite transparent: there is no allomorphy and no allophony of gender markers; gender agreement is always root-initial (and the proclitics in Table 7 are readily identifiable from their prosody, some of their segmental phonology, and the fact that they are separable, occurring in word-final positions when the verb is in second position). In Tsakhur it is quite opaque. There is a good deal of allomorphy, and this produces different patterns of syncretism: genders 1 and 4 syncretize in ‘hold’ but 1 and 2 in ‘hang’.<sup>6</sup> Gender is partly prefixal and partly infixal: infixal in formerly bipartite stems, where a former prefix has entrapped the root-initial gender

<sup>6</sup>In recent linguistic work on Daghestanian languages the genders are arbitrarily numbered.

marker, but the bipartite structure is ancient and not synchronically transparent. In both languages some but not all verbs take gender agreement: about 30% in Ingush and a very large majority in Tsakhur. Whether a verb takes agreement or not is then highly predictable for Tsakhur but much less predictable for Ingush; in this regard Ingush is less canonical.

In Tsakhur as in Ingush, the first two genders are used of humans and are referent-based, and the last two are arbitrarily assigned. In Avar (Nakh-Daghestanian; Daghestanian branch, Avar-Andic-Tsezic subbranch), gender is formally even simpler than in Ingush (in that for Avar there are no other verb prefixes and no proclitics, so gender markers are not just root-initial but word-initial) and entirely referent-based (there are three genders: masculine, feminine, and other, a.k.a. neuter). Also, unlike Ingush, the plural gender marker is entirely predictable from the singular one. The system is smaller than that of Ingush: three genders and four gender markers for Avar vs. eight genders and four markers for Ingush. The sole non-canonicity of Avar is that not all verbs and not all adjectives take gender agreement (about half of the verbs do, thus unpredictability is maximal).<sup>7</sup>

To summarize this section, non-canonicity can be a good guide to complexity and makes it possible to compare relative degrees of complexity using existing and straightforward criteria. Russian noun declension is considerably more complex than Mongolian; Tsakhur gender agreement is considerably more complex than that of Ingush or Avar; Ingush gender agreement is somewhat more complex than that of Avar. I have not attempted here a calculation of absolute complexity levels based on canonicity. (For a more detailed discussion of non-canonicity as complexity measure see Nichols 2016; forthcoming.)

---

<sup>7</sup>Avar is known for rampant multiple agreement in phrases and clauses: not only verbs and adjectives but also a number of adverbs, determiners, and other forms take agreement (Kibrik 1985; 2003). There are three possible analyses of multiple agreement in canonicity theory: (1) Gender is unnecessary, hence non-canonical in itself, so minimizing its use is canonical. (2) Multiple agreement is neutral, as long as all targets receive the same feature values (Corbett & Fedden 2016: 513) and agreement is obligatory (Corbett 2006: 14–15). (3) Given that gender exists, multiple agreement is canonical in that it demonstrates exhaustiveness of features across lexical classes (Corbett 2013: 54) and functional in that it increases consistency and identifiability of gender across different constituents and different utterances. I have no stance on this, but the sociolinguistic history of Avar may be relevant, as Avar is a spreading and inter-ethnic contact language of the type expected to undergo simplification (Trudgill 2011). In contrast, Ingush has undergone a poorly understood spread but is not an inter-ethnic or contact language, and Tsakhur is a small highland language and sociolinguistically quite isolated in Trudgill's sense (in which sociolinguistic isolation means no history of absorbing adult L2 learners; Tsakhur, like other highland Daghestanian languages, has very few adult L2 learners but is not at all isolated from contact of other kinds). If the spreading and inter-ethnic language has extensive multiple agreement, it may well be functional in some way, though canonicity and functionality are different things and not expected to coincide.

Table 8: Gender agreement in two Tsakhur verbs. Aorist tense. (Do-brushina 1999: 85 with some retranscription. *qq* = geminate, *y* = high back unrounded vowel, *X* = uvular.) Dot in citation form marks insertion point and boundary between the gender marker and the pieces of a bipartite stem. In actual inflected forms the gender marker has a dot on either side.

	<i>a.q-</i> ‘hold’	<i>giwa.X-</i> ‘hang’
1	<i>a.q.qy</i>	<i>giwa.r.Xyn</i>
2	<i>a.j.qy</i>	<i>giwa.r.Xyn</i>
3	<i>a.w.qu</i>	<i>giwa.p.Xyn</i>
4	<i>a.q.qy</i>	<i>giwa.t.Xyn</i>

### 3 Are gender languages more complex overall?

A possible explanation for the evolution of gender is that it arises easily, as some kind of excrescence or emergent category and probably due to reanalysis of existing markers, in a language that is already morphologically complex and already has at least some agreement as a model for gender agreement. And indeed, gender is almost never the sole inflectional category, or even just the sole agreement category.<sup>8</sup> If gender presupposes complexity, the synchronic result should be that when gender is disregarded languages with gender should still have higher overall complexity than languages without gender. To determine that, this section tests three hypotheses about the overall complexity of languages with and without gender. For all three I use the inventory complexity database of Nichols (2009), expanded to 196 languages with reasonably diverse genealogical and geographical distribution. It should be cautioned, though, that the database is slanted toward inflectional morphology of indexation and head marking, with better representation of categories such as person and classification than e.g. case or other categories of non-heads.<sup>9</sup>

<sup>8</sup> A possible exception is the western Nakh-Daghestanian languages, including Ingush and Avar discussed here, where there is no person agreement at all, but only gender agreement. (Arguably there is also number agreement, though that is usually treated as it is in Bantu languages, with number just a matter of gender pairing between singular and plural classes.)

<sup>9</sup> The reason for the imbalance is historical: the morphological measures are mostly drawn from the Autotyp database (Bickel et al. 2017), for which data on NP structure and noun inflection is a more recent addition and still incomplete. This is one reason why the database is best viewed as a convenience sample of categories than as a balanced sample of categories (much less an accurate measure of overall morphological complexity or even just overall complexity of inflectional morphology).

Hypothesis (i): Languages with gender are more complex overall than those without gender. For this count I used the entire set of complexity measures (phonological, morphological, syntactic, lexical), excluding gender; that is, measuring complexity other than in gender. The results are shown in Table 9: there is no significant difference in complexity between gender languages and genderless languages. What little correlation does show up is negative, contradicting the hypothesis.

Table 9: Overall complexity of languages with and without gender.

	Above	Below mean complexity	
Gender	28	38	
No gender	58	78	n.s. ( $p = 0.18$ ; Fisher 1-tailed)

Hypothesis (ii): Gender languages are more complex morphologically than genderless languages. This test uses the same survey except that only the morphological measures of complexity are counted. There is a significant negative correlation; see Table 10. Hypothesis (ii) fails, as does the null hypothesis; the finding here is that gender languages are less complex morphologically than genderless languages.<sup>10</sup>

Table 10: Overall morphological complexity of languages with and without gender. Figures in bold are above the expected values.

	Above	Below mean complexity	
Gender	15	<b>43</b>	
No gender	<b>60</b>	76	( $p = 0.01$ ; Fisher 1-tailed)

Hypothesis (iii): Gender languages have higher inflectional synthesis of the verb than genderless languages. Verb inflectional synthesis was defined as Categories per word (including roles) following the Autotyp database (Bickel et al. 2017). Again the hypothesis is falsified (Table 11).<sup>11</sup>

<sup>10</sup>But recall again the bias toward features of heads in the database, above in the text and note 9; to evaluate the impact of Table 10 it is especially important to have a balanced survey of categories.

<sup>11</sup>What small correlation emerges is negative. Bickel & Nichols (2013a) exclude role marking from verb synthesis; on that measure, there is a significant negative correlation, falsifying both survey and null hypotheses and suggesting that it is non-complexity that favors gender. Again (see notes 9 and 10) the result shows that a balanced morphological survey is important.



#### 4 *Why is gender so complex? Some typological considerations*

Table 11: Overall inflectional synthesis of the verb for languages with and without gender.

	Above	Below mean complexity	
Gender	22	36	
No gender	64	75	n.s. ( $p = 0.19$ , Fisher 1-tailed)

Thus, except for gender itself, on three criteria gender languages are no more complex than others and may even be less complex. The rise of gender must be due to something other than sheer complexity, and the synchrony of gender does not require or favor overall high complexity.

#### 4 Complexity in classifier systems: numeral classifiers, possessive classification

Perhaps systems of classification in general are complex, so that complexity is not just a peculiarity of gender. This section considers the complexity of numeral classifier and possessive classifier systems.

Numeral classifiers are well known from many East Asian languages, e.g. Mandarin. The systems tend to be large (50 or more in common use for Mandarin, plus many more that can be extracted from occasional occurrence in the long and varied written tradition of Chinese); the inventory complexity is therefore high. The numeral classifiers generally have independent phonological wordhood status and minimal or no sandhi, fusion, etc. and are semantically transparent, though with some flexibility as to what nouns take what classifiers (the flexibility is itself semantically motivated); therefore the descriptive complexity is low.

Elsewhere around the Pacific Rim numeral classifiers tend to be less transparent. Nivkh (isolate; Sakhalin Island and the lower Amur, eastern Siberia) has some 30 numeral classes (Mattissen 2003 gives the highest number) (moderate-high inventory complexity), in which the classifier is fused to the numeral, the combination being semi-transparent, and (at least in the recent and present situation of speech-community contraction and reduced functionality) different classifiers have different distributions: some classifiers apply only to the numerals 1–5, some to 1–5 and 10, and some to all of 1–10 (this is fairly high descriptive complexity). Yurok (Algic, northern California; Robins 1958: 86–91) has 15 classes (moderate inventory complexity), semantically motivated (human, plant, various shapes, etc.). The classifier is inextricably and opaquely fused with the numeral,

yielding a de facto system of 15 classes of numerals (high descriptive complexity). (“Several informants were aware of this complexity and would say admiringly of another speaker that he or she ‘knows the numbers’ or ‘can count in Indian’”: Robins 1958: 87n.)<sup>12</sup> The languages with numeral classifiers range from morphologically non-complex (Mandarin and other Southeast Asian languages) to morphologically complex (Yurok), with the major hotbed of numeral classifier systems found in the morphologically relatively simple languages of Southeast Asia but other languages with numeral classifiers sprinkled all around the Pacific Rim, where languages have high complexity in general. A preliminary conclusion is that numeral classifier systems can be complex in themselves but numeral classifier languages as a set are not more complex than others.

Possessive classes (Nichols & Bickel 2013; Bickel & Nichols 2013b) involve covert classification of nouns which becomes overt only when the noun has possessive morphology. Many languages have a distinction of two classes of nouns, usually termed alienable and inalienable. The formal difference can be as simple as obligatory possession of inalienables vs. optional possession of alienables, and the semantic opposition can be quite straightforward (e.g. kin terms and/or body parts vs. other nouns). In such a language (the most frequent type), both inventory and descriptive complexity are low. A complex system is that of Anêm (isolate, New Britain; Thurston 1982), in which possessed nouns fall into at least 20 classes marked by some simple and some composite suffixes and involving a mix of partly semantic and partly arbitrary classification (Thurston 1982: 37–38), very high inventory complexity. There is a good deal of syncretism between classes, and class membership is semantically unpredictable, so descriptive complexity is also high. The most complex system I have observed is that of Cayuvava (isolate, Bolivia; Key 1967), in which possessive morphemes are circumfixes with much allomorphy of both pieces and partial interdependence between the pieces. Both prefixal and suffixal parts appear to reflect person, and the suffixal part is also purely classificatory. The choice of classifier is semantically unpredictable. The set of first person singular forms is shown in Table 12. The inventory complexity is high and the descriptive complexity might be described as stratospheric.

Thus possessive classification, like numeral classification, can also be quite complex, and probably no less complex than gender. The overall complexity of

---

<sup>12</sup>Mattissen (2003) compiled the fullest list of Nivkh numeral classifiers by cross-tabulating lower figures reported in other sources. Robins compiled his list in analogous fashion from different speakers (“The table...was compiled from several informants and represents a collation of material from all of them, each accepting, though not necessarily volunteering, all the forms tabulated” [87]).

#### 4 Why is gender so complex? Some typological considerations

Table 12: First person singular possessive circumfixes in Cayuvava (Key 1967).

<i>a-</i>	...	<i>-i</i> <i>-ro</i> <i>-Ø</i> <i>-ai</i>
<i>i-</i>	...	<i>-i</i> <i>-Ø</i>
<i>ub-</i>	...	<i>-i</i>
<i>ku-</i>	...	<i>-i</i>
<i>či- ~ ič-</i>	...	<i>-i</i>
<i>č-</i>	...	<i>-ri</i>

languages with possessive classification ranges from low (as in Polynesian languages: see e.g. Wilson 1982 for Polynesian possessive classification) to high (e.g. Anêm, whose Austronesian-speaking neighbors consider it impossible to learn; Thurston 1982: 51).

Results of the same kinds of tests, for morphological complexity against presence vs. absence of numeral classifiers, possessive classes, or either one are shown in Tables 13–15. Again none of the results are significant: languages with classification of either type are not more complex than those without. There is, however, an interesting trend for a positive correlation of possessive classification and high complexity (Table 14), which merits testing on a larger sample.

Overall, then, neither gender, numeral classifiers, nor possessive classification appears to require or favor general morphological complexity as a diachronic prerequisite or synchronic correlate, and complex classification is not just a simple reflection of the overall complexity level of the language.

Table 13: Overall morphological complexity of languages with and without numeral classifiers

	Above	Below mean complexity	
Classifiers	14	16	
No classifiers	88	80	n.s. ( $p = 0.35$ ; Fisher 1-tailed)

Table 14: Overall morphological complexity of languages with and without possessive classification

	Above	Below mean complexity	
Poss. classes	38	45	
No poss. classes	41	74	n.s. ( $p = 0.099$ ; Fisher 1-tailed)

Table 15: Overall morphological complexity of languages with and without classification (numeral or possessive)

	Above	Below mean complexity	
Classification	41	45	
No classification	33	50	n.s. ( $p = 0.19$ ; Fisher 1-tailed)

## 5 Complexity in person indexation

Person, like gender, is primarily an agreement or indexation category, and in fact person is the clausal agreement category par excellence. Person indexation on verbs can be quite complex, and this section compares complexity and the evolution of complexity or non-complexity in gender and person systems, arguing that complex person marking systems can develop emergent alternative analyses that are simpler while gender systems do not and apparently cannot do this.

Inventory complexity of person marking is high in West Caucasian languages such as Adyghe and Abkhaz, which index six person-number categories for three roles, for an 18-cell total paradigm; Yimas (Lower Sepik-Ramu, New Guinea; Foley 1991) with 3 persons  $\times$  3 numbers  $\times$  2 roles (also 18), or Kiowa (Kiowa-Tanoan, U.S.; Watkins & McKenzie 1984), 3 persons  $\times$  3 numbers  $\times$  2 roles  $\times$  2 conjugation classes, plus direct/inverse marking for 17 subject-object paradigm cells (total of 53). In the West Caucasian languages transparency is high, since each argument is indexed by an unambiguous person-number marker in a separate slot, while transparency for Kiowa is low, since subject and object roles are indexed with mostly fused morphemes (see the paradigms in Watkins & McKenzie 1984: 115–116). The Kiowa non-transparencies and the two conjugation classes are non-canonical.

A different kind of non-canonicity is found in languages such as Laz (Kartvelian, Georgia and Turkey; Lacroix 2009: 283, Öztürk & Pöchtrager 2011: 48),

#### 4 Why is gender so complex? Some typological considerations

Table 16: Arhavi Laz subject and object indexation paradigm. Only one argument is overt. ... = root + thematic suffix. Phonological alternations not shown. (Lacroix 2009: 283, 298, plus examples on other pages; s.a. Öztürk & Pöchtrager 2011: 51.)

	S/A-	O- ...	-S/A	Examples
1SG	<i>b-</i>	<i>m-</i>		<i>b-dzir-om</i> 'I see him' <i>m-dzir-om</i> 'you <sub>sg</sub> see me'
2SG		<i>g-</i>		<i>dzir-om</i> 'you <sub>sg</sub> see him' <i>g-dzir-om</i>
3SG			<i>-s/n/u</i>	<i>dzir-om-s</i> 'he sees him' <i>m-dzir-om-s</i> 'he sees me'
1PL	<i>b-</i>			<i>b-dzir-om-t</i> 'we see him'
2PL		<i>g-</i>		<i>dzir-om-t</i> 'you <sub>pl</sub> see him'
3PL			<i>-an/nan/es/n</i>	<i>dzir-om-an</i> 'they see him'

where the two arguments of transitive verbs compete for a single person prefix slot and the competition is resolved by person and role hierarchies (1, 2 > 3, A > O). See Table 16, especially the first two forms listed, where the prefix is first person singular, subject in the first example *b-dzirom* and object in the second *m-dzirom*. The system is non-canonical in that the same slot can mark either subject or object, and in that second person has no overt marking at all. In addition to person/number prefixes, number is also indicated by a plural affix that registers plurality of any argument (A, S, O, G) if it is first or second person, and another that indexes number for a third person S/A.<sup>13</sup> This is non-canonical in that a single category (plural) is marked with different formatives that have different distributions (third person subject indexation vs. non-third-person plural argument registration).

The argument indexation system of Tundra Yukagir (isolate, Siberia; Maslova 2003b) is even less canonical; see Table 17. The system is a proximate/obviative one somewhat like those of Tagalog, Algonquian languages, and others (see Bickel

<sup>13</sup>I use *index* and *register* as in Nichols (1992: 48–49): indexation copies or otherwise marks features of the argument (person, number, etc.) on the verb, while registration simply indicates the presence of an argument in the clause but does not agree with or copy features. I assume that what is called promiscuous number marking (Leer 1991) is not indexation (of number on an argument marker, because the argument is not specified) but registration (of a multiple argument, a category similar to pluractionality and easily overlapping with it: see Wood 2007, Yu 2003).

Table 17: Tundra Yukagir obviation system (Maslova 2003b: 18). Focus = proximate. S focus column constructed from other tables in Maslova (2003b) and Kolyma Yukagir (Maslova 2003a).

	Neutral transitive	O focus	A focus intransitive	Neutral	S focus
1SG	-Ø-ng	-me-ng	-Ø	-je-ng	-l
2SG	-me-k	-me-ng	-Ø	-je-k	-l
3SG	-m-Ø	-me-le	-Ø	-j-Ø	-l
1PL	-j	-l	-Ø	-je-l'i	-l
2PL	-mk	-mk	-Ø	-je-mut	-l
3PL	-nga	-ngu-me-le	-ngu-Ø	-ngi	-ngu-l

2011 for this typology), in which one of the arguments is designated as proximate (usually because of topicality or a similar parameter) and the others are obviative. Verb indexation and noun case track proximate and obviative status. (The term for 'proximate' in Tagalog and Yukagir descriptions is usually *focus*.) In Yukagir, unlike other languages with obviation, a proximate argument is not required, and unlike Tagalog the proximate argument can be only A, S, or O (for Kolyma Yukagir, only A or S; Maslova 2003a). Identifying single-function forms that index person/number categories is impossible for most of the cells. Nearly every cell in Table 17 exhibits one or more non-canonicalities.

To judge from the languages surveyed here, person systems can have greater inventory complexity and greater descriptive complexity (more non-canonicalities) than gender systems. However, person systems also have simpler and more canonical analyses available than gender systems do: hierarchical structuring, in which different patterns that violate biuniqueness reduce to a single ordering principle. The Laz paradigm shown in Table 16 reduces to a set of signs plus two hierarchical patterns: 1, 2 > 3 and A > O (for discussion of the Pazar Laz hierarchies see Öztürk & Pöchtrager 2011: 48). Maslova (2003b: 17, 20) reduces much of the complexity and non-transparency of Table 17 to the two hierarchies illustrated in Tables 18 and 19 and summarized in Table 20.

On this perspective, the Yukagir system is still less than straightforward, and it differs from better-known obviation systems in that it tracks the proximate/obviative status of the O while the others mainly track the A. But the individual morphemes are better motivated and the whole system emerges as less non-canonical than the non-hierarchical one, and thus as less complex.

#### 4 Why is gender so complex? Some typological considerations

Table 18: Tundra Yukagir obviation: Distribution of transitive markers (Maslova 2003b: 17). Bracketed comment mine.

Person of A:	A focus	Neutral	O focus
1	-Ø-	-Ø-	- <i>me</i> -
1+ other [ <i>i.e.</i> 1PL]	-Ø-	- <i>j</i>	- <i>l</i>
Non-1	-Ø-		- <i>m(e)</i> -

Hierarchy: Focus > Speaker > other. Zero suffix signals that A outranks O in this hierarchy.

Table 19: Tundra Yukagir obviation: Person slot (the second element of the internally hyphenated forms in Table 17) in the O focus paradigm (Maslova 2003b: 20)

	O neutral	O focus	
1SG	- <i>ng</i>	- <i>ng</i>	A = SAP
2SG	- <i>k</i>	- <i>ng</i>	A = SAP
2PL	- <i>k</i>	- <i>k</i>	A = 2 + 3 [ <i>i.e.</i> 2PL]

Hierarchy: SAP > other

Table 20: Summary of hierarchical effects in Tundra Yukagir obviation. (Recall that Focus = proximate.)

Hierarchy	What it determines
Obviation: Focus > Speaker > other	Form of person/number markers
Role: A > O	Zero vs. nonzero suffix
Person: SAP > other	Form of second slot in person/number marker

All forms index the A (relying on hierarchies) and register an O.

Hierarchy for access to O registration: Focus > all else.

A striking example comes from Alutor (Chukchi-Kamchatkan). Paradigms, too long to reproduce here, for the most basic forms are in Nagayama (2003), Mal'ceva (1998), and others; full tables are in Kibrik et al. (2004: 639–648). The tables are not only long but complex and with dauntingly little correlation of form to function, either within or across paradigms. Kibrik (2003) reduces the forms to a basic person hierarchy of 1SG, 1PL, 2SG > 2PL, 3 for access to the A slot, the reverse for access to O, for relatively polar A and O (and additional provisions for less polar A and O), plus different cutoffs in different mood categories based in part on the speaker's control over, or ability to predict, the event.

Hierarchically based indexation (in which I also include inverse indexation) has the advantage that less information is required than for standard paradigm-based accounts. Roles and/or person can be inferred from hierarchies rather than being fully specified. Those hierarchies are not part of the description of each paradigm; they are grammar-wide, to some extent even universal, as are cross-linguistically favored cutoff points such as 1, 2 > 3 person or S/A > O. For purposes of assessing descriptive complexity, a grammar-wide principle does not have to be specified for particular paradigms and adds no information to their description; a universal principle does not contribute information to any particular grammar.

In these respects, hierarchical indexation may well be canonical. Viewed in the proper perspective, it is not a type of paradigm but what might be called a BLUEPRINT for creating paradigms and forms. Henceforth I will use the term BLUEPRINT because it is not a precise theoretical term and because it implies an instruction or algorithm or the like rather than a structure or set of forms. (How to implement hierarchical and other blueprints in theoretical morphology is a challenge not addressed here.) The paradigm is the blueprint's output, and available evidence indicates that describing the output requires more information than describing the blueprint.

A cross-linguistically recurrent minimal hierarchical system shows up in verbs indexing two arguments, where combinations of first and second person ('I VERB you', 'you VERB me') are often opaque, or overtly mark only one of the persons, or are ambiguous or otherwise non-transparent (Heath 1991; 1998). This amounts to treating the participant scenario not as a pair of arguments and not even as a morphologically fused dyad but as a monad. From what is left unarticulated, plus culture-specific and universal expectations, one can infer who does what to whom; see Heath's detailed analysis. This too is a type of blueprint.

The theoretical claim of Kibrik (2003: 376) for Alutor is that identical forms point to proximity in cognitive space, and the structure of that space is much



less complex than traditional conjugation tables. This statement, and other descriptions of hierarchies, strike me as presenting a view of an alternate, simpler paradigm, but nonetheless a paradigm and not a blueprint.

Person differs from gender and other agreement and classification categories in that only person exhibits hierarchical patterning. Gender and classifiers never do, in my experience. Even in the concurrent gender and classifier system of Mian described by Corbett & Fedden (2016), where one might expect the two systems to compete for a single slot at least in some circumstances, this does not happen. Number and gender can of course be drawn into the patterning of person if they are drawn along in coexponential markers, but on their own they do not form hierarchies.

The reason for this may lie in the fact that person markers are typically, perhaps always, referential. There are three views on whether person markers are referential. One view is that person markers are always referential, not only the pronominal arguments of pro-drop languages but also the person agreement affixes of languages like English or German or Russian, where there is generally a clearly referential overt argument as well as the verbal person marker whose referentiality is at issue (Kibrik 2011). The second view is that person markers are never referential, even in pro-drop languages, but reference arises from the context and the arguments and is attributed to markers in processing or grammatical analysis (Evans 1999; 2003). The third view is that some person markers are referential and some are not: those variously described as pronominal arguments or cross-reference are referential while those described as agreement are not referential but are simply categories of referring NPs (Hengeveld 2012). Whichever view one adopts, it is probably safe to say that if anything is referential in verb indexation, person is. That is, in proneness to referentiality, person > other categories.

I doubt that categories other than person are ever referential. Gender, in particular, appears to never be referential.<sup>14</sup> Creissels (2014) shows that verbs in Avar (Nakh-Daghestanian, eastern Caucasus) are entirely ambiguous between anaphoric, unspecified, and absent readings of one or more arguments. (1) gives examples parallel to his from Ingush, where the grammar is identical in this respect. Ingush can be described as having two zero pronominals, one anaphoric

---

<sup>14</sup> I use *referential* of gender in the same way as I used it of person in the previous paragraph, so that *is referential* means 'refers' or 'can refer'. This is the usage of Kibrik (2011). It is not to be confused with the same word in Dahl's distinction (2000) of referential gender (= my referent-based gender) vs. lexical gender. Both senses of the word are established in the literature; I chose the one having to do with a new point made here, though Dahl's term is probably the earlier one. The issue needs to be resolved; my *referent-based* is only a patch.

and one unspecified, and the first two readings have these as A argument. The third reading has no A at all; this kind of clause, in which the A is absent but the O remains an O and is not promoted to S, is not found as a major clause type in European languages.<sup>15</sup> (2) shows that exactly the same readings are available to a verb that does not take gender agreement (recall from above that gender is a partial category in Ingush). This shows that gender has nothing to do with referentiality in Ingush. (No argument can be made for either Ingush or Avar about referentiality of person, as both languages lack an inflectional category of person.)

(1) Ingush<sup>16</sup>

a. Anaphoric zero:

Ø yz v.iira

X<sub>i</sub> 3SG V.killed

‘(I/you/he/she/they) killed him.’

b. Unspecified zero:

Ø yz v.iira

UNSP 3SG V.killed

He was killed (by someone);

(Someone) killed him;

‘They killed him.’

c. Absent A:

yz v.iira

3SG V.killed

‘He was/got killed.’

(2) Ingush

a. Anaphoric zero:

Ø yz leacar

X<sub>i</sub> 3SG V.caught

‘(I/you/he/she/they) caught him.’

---

<sup>15</sup>It is not that this verb has ambitransitive (labile) valence; in Ingush this construction seems to be available to all transitive verbs and perhaps all two-argument verbs more generally. Actual ambitransitive valence of the type (A)O occurs in very few Ingush verbs (I know of only the five listed in Nichols 2011: 466–467).

<sup>16</sup>All verbs in (1)–(2) are in the witnessed past tense (a.k.a. aorist). The nonwitnessed tense (*v.iina.v*, *leacaa.v*), which is resultative and/or inferential evidential, would probably be more likely for the (c) examples.

#### 4 *Why is gender so complex? Some typological considerations*

b. Unspecified zero:

Ø yz leacar

UNSP 3sg V.caught

‘He was caught (by someone)’;

‘(Someone) caught him’;

‘They caught him.’

c. Absent A:

yz leacar

3sg V.caught

‘He was caught/arrested.’

All reviewers of this chapter, and most audiences where I have presented this part of it, raise the objection that gender is referential: it is referential in English pronouns, and gender is known to be important in reference tracking. The point merits a brief excursus. As background, saying that a morpheme or category is referential means that it refers, or carries reference, or bears a referential index. If a category is referential, the category itself is what refers, and not the word that carries that category. English pronouns certainly refer, but it is the pronoun and not its gender that is referential. English pronouns are no more (and no less) referential than those of e.g. Finnish or Turkish (languages which have no gender in either nouns or pronouns) or Ingush (which has noun gender but no pronoun gender), or for that matter French or Russian (which have gender in nouns and pronouns). The presence or absence of gender in pronouns, or whether the gender (in languages that have it) is entirely natural (as in English) or agrees with a noun antecedent (as in French or Russian), does not affect the referentiality of pronouns.

Gender has indeed often been said to be useful in reference tracking, but in fact its usefulness in this function is marginal, as human protagonists of narrative and discourse often belong to the same gender. Kibrik (2011: 334–360) makes this claim and supports it with cross-linguistic, discourse, and experimental evidence, and also emphasizes that reference tracking is not the same as referring: reference tracking mostly has to do with disambiguating and resolving potential referential conflicts.

To summarize on referentiality, person can be referential, and perhaps person is always, and necessarily, referential; but gender is not referential.<sup>17</sup> Numeral

<sup>17</sup>My own strong intuition is that inflected verb forms in Ingush do not refer. A context like the anaphoric one in (1a)–(2a) can make it unambiguous who performed and underwent the action, and the choice of witnessed vs. non-witnessed evidentiality categories can make clear whether the speaker knows who did what, but the verb form itself does not refer and the gender at most guides the search for an antecedent by narrowing down its possible gender.

classifiers and possessive classifiers are probably also not referential, but as they appear in NPs rather than on verbs the question of referentiality is less clear.<sup>18</sup>

I am not aware that the matter has been the subject of research, but I suggest a diachronic scenario like the following. On verbs that index two arguments, and especially when person agreement develops enough complexity and/or opacity (e.g. in fusion of forms), hierarchical patterns can arise. The most likely first step occurs when phonological change has made formerly discrete A and O person markers opaque and universal person hierarchies step in to disambiguate, and in doing so they impose their own order. Hierarchical structure is thus an emergent pattern, and it functions not in the usual way that paradigms and sets of forms do but in a new way, as a blueprint. A blueprint is functional where complexity is high, because it reduces the complexity. The ability to function referentially seems to be critical to this emergence, perhaps because referentiality makes it possible to draw on universal hierarchies and fix 1<>2 person forms as morphologically opaque monads.<sup>19</sup>

The reason why gender systems can be so complex is then that they have no self-correcting mechanism like the hierarchical blueprint that might simplify them, and they are stable enough that complexity can build up over time without causing the whole system to be shed. Not only are they stable within families; the complex interaction of gender with case and number persisted in Latin, ancient Greek, late Proto-Slavic, and early Germanic, despite large spreads with absorption of substantial numbers of L2 learners, circumstances that are expected to simplify languages but did not appreciably simplify the paradigms of these languages.

The papers by Liljegren (2019 [this volume]) and Garbo & Miestamo (2019 [in Volume II]) (and also Maho 1999) show examples of gender systems simplifying, but the way in which they go about simplifying supports my point. Both papers describe changes in which closer alignment of semantics and gender classification occurs in individual words, beginning with a few words and at the extreme ends of Corbett's agreement hierarchy (1991: 248–259). Typically, a word referring to an animate or human but with an arbitrary gender classification begins to trigger an appropriate animate or human gender agreement marker in limited contexts (such as predicate nominal). Over time, more words and more contexts are involved, and eventually the system ends up based on animacy rather than on

---

<sup>18</sup>Numeral classifiers can fuse to demonstratives and those can be referential and can furthermore be accreted to verbs as indexes, but by that point they have begun to function as third person markers which also index classificatory categories.

<sup>19</sup>1<>2 is Heath's now widely used notation for opaque morphemes that are ambiguously 1>2 and 2>1 (1998).

arbitrary classifications. The early stages, however, add complexity, as the gender agreement rules refer to contexts, create alternations and options for some words but not others, and otherwise introduce variation. Alternatively, gender can be lost when gender agreement is lost, and in the languages Di Garbo and Miestamo study, where singular and plural nouns mostly have different gender agreement markers and gender is marked not only in agreement but also on the nouns themselves, the former gender marking changes into a system of number marking. But these are all developments where gender is ultimately simplified by reduction or loss, while I am talking about complex person systems which retain all their categories and markers but in some kind of reanalysis acquire an emergent alternative analysis as blueprint-driven. For this, I believe, we have no analog in gender.

## 6 Stability of gender

Gender is very stable in language families (Matasović 2007; 2014). In Indo-European, gender – the categories, the markers, and the complex interaction with case paradigms – lasts as long as the original case endings do, so the original system is still largely in place in Baltic and Slavic and to some extent in Germanic (where parts of it are recognizable to the specialist). More precisely, gender does not outlast the original case endings – nor, usually, vice versa (though Armenian is a counterexample: see Kulikov 2006). Even when case was lost in the various Romance languages and in Macedonian and Bulgarian, the gender categories have remained and their markers continue those of early Indo-European. Whatever the reason for this stability, it means that a gender system can evolve considerable complexity without much risk that the language will abandon it or restructure it. The complexity of the Slavic gender system is simplified not by restructuring but by losing case entirely, in Macedonian and Bulgarian; this removes all the complexity that is due to cumulative expression of case with gender, discussed in §2 above. In general in Indo-European, where gender has been lost, case has generally also been lost, as in English or some Iranian languages (e.g. Persian). Loss of gender has happened in three languages and one additional dialect of Nakh-Daghestanian, a very old family (probably older than Indo-European) with about 40 daughter languages, so 10% or less of the family has lost gender. In these languages gender is not cumulative with case but is expressed only in agreement, and languages that lose gender keep case. The languages that have lost gender have histories of large spreads and contact of the kind expected to simplify languages; but not all of the languages with similar sociolinguistic histories have lost gender. The prehistory of gender in Nakh-Daghestanian is still

poorly understood (though see Schulze 1998), but the complexity of gender marking in Tsakhur, discussed above, is a clearly secondary phenomenon caused by positional sound changes after the accretion of spatial prefixes entrapped the gender prefixes. Some high-contact languages have reduced the number of gender markers and categories, but gender is retained and the agreement rules function in much the same way across the family.

Neither the inventory and descriptive complexity of Nakh-Daghestanian gender, nor the descriptive complexity of conservative Indo-European languages, nor any other gender system I am aware of, has any self-correcting mechanism like hierarchical patterning for person.

## Acknowledgments

I thank the editors and four anonymous referees for extremely helpful comments. Research on Ingush was supported by NSF 96-16448 and Ingush fieldwork was carried out in the Max Planck Institute for Evolutionary Anthropology, Leipzig, from 2002 to 2014. Analysis and writing were supported by a grant from the Russian Academic Excellence Project 5-100 to the Higher School of Economics, Moscow.

## References

- Audring, Jenny. 2017. Calibrating complexity: How complex is a gender system? *Language Sciences* 60. 53–68.
- Bickel, Balthasar. 2011. Grammatical relations typology. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 399–444. Oxford: Oxford University Press.
- Bickel, Balthasar & Johanna Nichols. 2007. Inflectional morphology. In Timothy Shopen (ed.), *Language typology and syntactic description*, 2nd edn., vol. 3, 169–240. Cambridge: Cambridge University Press.
- Bickel, Balthasar & Johanna Nichols. 2013a. Inflectional synthesis of the verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/22>.
- Bickel, Balthasar & Johanna Nichols. 2013b. Obligatory possessive inflection. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/58>.

#### 4 Why is gender so complex? Some typological considerations

- Bickel, Balthasar, Johanna Nichols, Taras Zakharko, Alena Witzlack-Makarevich, Kristine Hildebrandt, Michael Rießler, Lennart Bierkandt, Fernando Zúñiga & John B. Lowe. 2017. *The AUTOTYP typological databases*. <https://github.com/autotyp/autotyp-data/tree/0.1.0>. Version 0.1.0.
- Corbett, Greville G. 1982. Gender in Russian: An account of gender specification and its relation to declension. *Russian Linguistics* 2. 197–232.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2007. Canonical typology, suppletion, and possible words. *Language* 83(1). 8–42.
- Corbett, Greville G. 2013. Canonical morphosyntactic features. In Dunstan Brown, Marina Chumakina & Greville G. Corbett (eds.), *Canonical morphology and syntax*, 48–65. Oxford: Oxford University Press.
- Corbett, Greville G. 2015. Morphosyntactic complexity: A typology of lexical splits. *Language* 91(1). 145–193.
- Corbett, Greville G. & Sebastian Fedden. 2016. Canonical gender. *Journal of Linguistics* 52(3). 495–531.
- Creissels, Denis. 2014. P-lability and radical P alignment. *Linguistics* 52(4). 911–944.
- Dahl, Östen. 2000. Animacy and the notion of semantic gender. In Barbara Unterbeck (ed.), *Gender in grammar and cognition*. Vol. 1: *Animacy and the notion of semantic gender: Approaches to gender*, 99–115. Berlin: Mouton de Gruyter.
- Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
- Dobrushina, Nina. 1999. Glagol. In Alexander E. Kibrik (ed.), *Èlementy caxurskogo jazyka v tipologičeskom osvesčeenii*, 58–85. Moscow: MGU.
- Donohue, Mark & Johanna Nichols. 2011. Does phoneme inventory size correlate with population size? *Linguistic Typology* 15(2). 161–170.
- Evans, Nicholas. 1999. Why argument affixes in polysynthetic languages are not pronouns: Evidence from Bininj Gun-Wok. *Sprachtypologie und Universalienforschung* 52(3/4). 255–281.
- Evans, Nicholas. 2003. *Bininj Gun-Wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: Pacific Linguistics.
- Foley, William A. 1991. *The Yimas language of New Guinea*. Stanford: Stanford University Press.
- Garbo, Francesca Di & Matti Miestamo. 2019. The evolving complexity of gender agreement systems. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume II: World-wide comparative studies*, 15–60. Berlin: Language Science Press. DOI:??

- Heath, Jeffrey. 1991. Pragmatic disguise in pronominal-affix paradigms. In Frans Plank (ed.), *Paradigms: The economy of inflection*, 75–89. Berlin: Mouton De Gruyter.
- Heath, Jeffrey. 1998. Pragmatic skewing in 1<-->2 pronominal combinations in Native American languages. *International Journal of American Linguistics* 64(2). 83–104.
- Hengeveld, Kees. 2012. Referential markers and agreement markers in Functional Discourse Grammar. *Language Sciences* 34(4). 468–479.
- Janhunen, Juha. 2012. *Mongolian*. Amsterdam: John Benjamins.
- Key, Harold H. 1967. *Morphology of Cayuvava*. The Hague: Mouton.
- Kibrik, Aleksandr E., Sandro V. Kodzasov & Irina A. Muravyova. 2004. *Language and Folklore of the Alutor people* (Endangered Languages of the Pacific Rim A2-042). Osaka: ELPR.
- Kibrik, Alexander E. 1985. Toward a typology of ergativity [selections]. In Johanna Nichols & Anthony C. Woodbury (eds.), *Grammar inside and outside the clause: Some approaches to theory from the field*, 268–323. Cambridge: Cambridge University Press.
- Kibrik, Alexander E. 2003. *Opyt morfoložičeskoj rekonstrukcii kognitivnoj struktury (na materiale sfery ličnogo dejksisa v aljutorskom jazyke)*. In: Alexander E. Kibrik: *Konstanty i peremennye jazyka*, 369–377. St. Petersburg: Aleteija.
- Kibrik, Andrej A. 2011. *Reference in discourse*. Oxford: Oxford University Press.
- Kulikov, Leonid. 2006. Case systems in a diachronic perspective: A typological sketch. In Leonid Kulikov, Andrej L. Malchukov & Peter de Swart (eds.), *Case, valency and transitivity*, 23–47. Amsterdam: John Benjamins.
- Lacroix, René. 2009. *Description du dialecte laze d'Arhavi (caucasique du sud, Turquie): Grammaire et textes*. Université Lumière Lyon 2 dissertation.
- Leer, Jeff. 1991. Evidence for a Northern Northwest Coast language area: Promiscuous number marking and periphrastic possessive constructions in Haida, Eyak, and Aleut. *International Journal of American Linguistics* 57(2). 158–193.
- Liljegren, Henrik. 2019. Gender typology and gender (in)stability in Hindu Kush Indo-Aryan languages. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 279–328. Berlin: Language Science Press. DOI:??
- Maho, Jouni. 1999. *A comparative study of Bantu noun classes* (Orientalia et Africana Gothoburgensia 13). Göteborg: Acta universitatis gothoburgensis.
- Mal'ceva, Alla A. 1998. *Morfologija glagola v aljutorskom jazyke*. Novosibirsk: Sibirskij xronograf.
- Maslova, Elena. 2003a. *A grammar of Kolyma Yukaghir*. Berlin: Mouton de Gruyter.



#### 4 Why is gender so complex? Some typological considerations

- Maslova, Elena. 2003b. *Tundra Yukaghir*. Munich: Lincom Europa.
- Matasović, Ranko. 2014. Verbal and nominal agreement: Areal distribution and typological correlations. *Linguistic Typology* 18(2). 171–214.
- Matasović, Ranko. 2007. The diachronic stability of gender revisited. Presented at Association for Linguistic Typology 8th Biennial Meeting, Paris.
- Mattissen, Johanna. 2003. *Dependent-head synthesis in Nivkh: A contribution to a typology of polysynthesis*. Amsterdam: John Benjamins.
- 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.
- Nagayama, Yukari. 2003. *Ocherk grammatiki aljutorskogo jazyka* [Grammatical Outline of Alutor]. (Endangered Languages of the Pacific Rim A2-038). Osaka: ELPR.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Nichols, Johanna. 2009. Linguistic complexity: A comprehensive definition and survey. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an eng variable*, vol. i, 110–125. Oxford: Oxford University Press.
- Nichols, Johanna. 2011. *Ingush grammar*. Berkeley: University of California Press.
- Nichols, Johanna. 2016. Complexity as non-canonicity: An affordable metric for morphology. SLE Annual Meeting, Leiden. (Submitted for Peter Arkadiev and Francesco Gardani, eds., *The complexities of morphology*.)
- Nichols, Johanna. forthcoming. Canonical complexity. In Peter Arkadiev & Francesco Gardani (eds.), *The complexities of morphology*. Oxford: Oxford University Press.
- Nichols, Johanna & Balthasar Bickel. 2013. Possessive classification. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/59>.
- Öztürk, Balkız & Markus A. Pöchtrager (eds.). 2011. *Pazar Laz*. Munich: Lincom Europa.
- Robins, Robert H. 1958. *The Yurok language, grammar, texts, lexicon* (University of California Publications in Linguistics 15). Berkeley: University of California Press.
- Schulze, Wolfgang. 1998. *Person, Klasse, Kongruenz. Fragmente einer Kategorialtypologie des einfachen Satzes in den ostkaukasischen Sprachen*. Vol. 1: *Die Grundlagen (1. Teil)*. Munich: Lincom Europa.
- Shosted, Ryan K. 2006. Correlating complexity: A typological approach. *Linguistic Typology* 10(1). 1–40.

- Svantesson, Jan-Olof. 2003. Khalkha. In Juha Janhunen (ed.), *The Mongolic languages*, 154–176. London: Routledge.
- Thurston, William R. 1982. *A comparative study in Anêm and Lusi*. Canberra: Australian National University.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. New York: Oxford University Press.
- Watkins, Laurel J. & Parker McKenzie. 1984. *A grammar of Kiowa*. Lincoln: University of Nebraska Press.
- Wilson, William H. 1982. *Proto-Polynesian possessive marking*. Canberra: Australian National University.
- Wood, Esther J. 2007. *The semantic typology of pluractionality*. University of California, Berkeley dissertation.
- Yu, Alan C-L. 2003. Pluractionality in Chechen. *Natural Language Semantics* 11(3). 289–321.

**Part II**

**Africa**



## Chapter 5

# Niger-Congo “noun classes” conflate gender with deriflection

Tom Güldemann

Humboldt University Berlin and Max Planck Institute for the Science of Human History

Ines Fiedler

Humboldt University Berlin

This paper reviews the treatment of gender systems in Niger-Congo languages. Our discussion is based on a consistent methodological approach, to be presented in §1, which employs four analytical concepts, namely agreement class, gender, nominal form class, and deriflection and which, as we argue, are applicable within Niger-Congo and beyond. Due to the strong bias toward the reconstruction of Bantu and wider Benue-Congo, Niger-Congo gender systems tend to be analyzed by means of a philologically biased and partly inadequate approach that is outlined in §2. This framework assumes in particular a consistent alliterative one-to-one mapping of agreement and nominal form classes conflated under the philological concept of “noun class”. One result of this is that gender systems are recurrently deduced merely from the number-mapping of nominal form classes in the nominal deriflection system rather than from the agreement behavior of noun lexemes. We show, however, that gender and deriflection systems are in principle different, illustrating this in §3 with data from such Niger-Congo subgroups as Potou-Akanic and Ghana-Togo-Mountain. Our conclusions given in §4 are not only relevant for the historical-comparative and typological assessment of Niger-Congo systems but also for the general approach to grammatical gender.

**Keywords:** gender, Niger-Congo languages, agreement, noun classes, deriflection.



## 1 The cross-linguistic approach to gender

Gender is understood here in terms of Corbett (1991), namely as systems of nominal classification (also called categorization) that are reflected by agreement. “With about two thirds of all African languages [being] gender languages” (Heine 1982: 190), Africa is rightly identified by Nichols (1992: 131) as a global hotbed of this phenomenon. At the same time, the majority of African languages belong to a single language family, Niger-Congo,<sup>1</sup> which displays a cross-linguistically unusual type of nominal classification described in a particular philological tradition. The existing research bias toward this large family keeps influencing the treatment of noun classification not only in African linguistics but also in typology in general. This contribution approaches the typical gender systems of Niger-Congo from a cross-linguistic perspective by subjecting them to an analysis that is universally applicable rather than one that is biased toward the special characteristics of this language group.

As mentioned above, according to the typologically most widespread approach, gender is the intersection of two domains, namely nominal classification and syntactic agreement, as the overt expression of a feature of a “trigger” (also called controller), usually a noun, on another word as the “target”. Several complications for the analysis of gender arise from Corbett’s (2006) extensive cross-linguistic survey of agreement. Notably, a language may have more than one agreement system and, more importantly for our discussion, a system sensitive to gender need not be restricted to this feature but most often also concerns others like number, person, case, etc. The features that a noun trigger transfers to a target not only relate to properties of an abstract lexical item, which are recurrently semantic. They can also concern the formal properties of the concrete word form of a given noun in the agreement context. A sound understanding of a gender system thus presupposes an exhaustive analysis of the language’s agreement system regarding all its agreement features and the subsequent “subtraction” of all factors but gender. If gender is only conflated with number, which is cross-linguistically frequent, it can be conceptualized as “agreement minus number.” This also holds for the Niger-Congo systems at issue here.

---

<sup>1</sup>We will not deal here with the still controversial question of the exact composition of this language family. That there is a substantial core group of genealogically related languages has been shown by Westermann (1935) with reference to gender, the very feature at issue, and the present discussion is concerned with languages that are robust members of this lineage (see Güldemann (2018) for a detailed recent discussion of the genealogical classification of African languages and the status of Niger-Congo in particular). While the discussion is also relevant for uncertain members of the group, we will not deal with them here.

The present contribution provides a novel analytical approach to gender. That is, we apply a strict distinction of four concepts, which are necessary whenever gender is reflected by syntactic agreement as well as nominal morpho-phonology, the latter implying some amount of what Corbett (1991) calls formal class assignment. The four notions are:<sup>2</sup>

- a. AGREEMENT CLASS (to be abbreviated as AGR and numbered by Arabic numbers),
- b. GENDER (to be occasionally labeled semantically or numbered by Roman numbers),
- c. NOMINAL FORM CLASS (to be abbreviated as NF and represented by the capitalized exponent), and
- d. DERIFLECTION (see p. 99 for the definition of the term, to be represented by the relevant NF set).

This approach is illustrated with the following example from the Bantu language Swahili, where agreement and nominal form classes are bold-faced in both vernacular and annotation line.

(1) Swahili (personal knowledge)

- a. ***m-toto***        ***yu-le***    ***m-moja*** ***a-me-anguka***  
**M(w)**-child(1) 1-D.DEM 1-one    1-PERF-fall  
‘that one child has fallen’
- b. ***wa-toto***        ***wa-le***    ***wa-wili*** ***wa-me-anguka***  
**w(A)**-child(2) 2-D.DEM 2-two    2-PERF-fall  
‘those two children have fallen’

The subject nouns in (1) trigger agreement on three targets: the demonstrative modifier *-le*, the numeral modifiers *-moja* and *-wili*, and the verb *-anguka* in the form of subject cross-reference. There are two different AGREEMENT CLASSES, AGR1 and AGR2, that are associated with the noun forms *m.toto* ‘child (SG)’ in (1a) and *wa.toto* ‘children (PL)’ in (1b), respectively, and they are evident from two different sets of exponents across the three relevant agreement targets, namely *yu-/m-/a-* vs. *wa-/wa-/wa-*. An agreement class in the present conceptualization

---

<sup>2</sup>Since genders and deriflections also establish sets of nouns, they could also be called “gender CLASSES” and “deriflection CLASSES”, respectively. We use here the short versions.

is thus a set of noun forms that share an identical behavior across all agreement contexts of a given system and thus equals what Corbett (1991, 2006) calls a “consistent agreement pattern” (see this author’s detailed discussion of the possible problems in establishing such an agreement class). (For schematic presentation, an agreement class is represented conventionally by the set of exponents of a single agreement target that involves the maximal class differentiation.) A crucial feature of our approach is that it is of no concern whether noun forms of one agreement class are of the same gender, number or any other feature, which differs from Corbett’s approach inspired by Zaliznjak (1964). An agreement class in the present terms is thus an overt but normally conflated reflex of diverse grammatical features – in Swahili, concretely of gender and number (see below for more details about our analytical and terminological differences to Corbett’s approach).

GENDER (CLASSES) are defined in line with Corbett’s (1991) cross-linguistic approach. Analytically, they are derived by abstracting from all other agreement features, which in the Swahili system is only number. The majority of Swahili nouns have a singular and a plural form so that a gender is instantiated by a particular pairing of the respective agreement classes. In (1), these are singular AGR1 and plural AGR2, which is the regular agreement behavior for count nouns of the “human” gender, which includes the nominal lexeme *-toto* ‘child’. The gender of transnumeral<sup>3</sup> nouns outside the systems of number distinctions is accordingly discernible from a single agreement class.<sup>4</sup> Normally, genders as the ultimate goal of analysis here are thus classes of nouns in the lexicon. However, gender often transcends the lexicon and applies to a language’s reference world more generally. That is, relevant systems can entail in addition such phenomena as nominal derivation and even the expression of grammatical relations. Swahili, for instance, also has agreement patterns (and noun prefixes) for derivational diminutives, infinitives, and various locative notions. The nominal lexeme *-toto*

---

<sup>3</sup>The term “transnumeral” is used here neutrally to refer to nouns that do not partake in the normal number oppositions of a language. It must not be confused with “general number” in terms of Corbett (2000: 9–19), which refers to a feature value in the number system as opposed to the more common singular and plural. Typically, transnumeral nouns like infinitives, locatives and non-count nouns for masses, liquids, abstracts etc. do not have different number forms, while general number is a number value that applies to nouns that have an alternative singular and/or plural variant.

<sup>4</sup>In general, any agreement class that only encodes gender and no other agreement feature does not require a distinction between gender and agreement class. An entire system of this kind would represent “ideal” functionally transparent gender marking, because there is a straightforward relation between one form and one meaning. However, such cases turn out to be rare cross-linguistically; they are found, for example, in Australian languages.



‘child’, for example, can also occur in the gender AGR7/AGR8 for diminutives, then appearing accordingly as *ki-toto/vi-toto* ‘baby/babies’.

Example (1) also shows the intimate interaction between nominal morphology and gender in Swahili. The subject nouns as the agreement triggers again exhibit two morphologically distinct word forms rendered by prefixes, namely *m-* and *wa-*, which characterize NF *M(W)-* and NF *W(A)-*, respectively. This direct morphological reflex of gender on the noun is conventionally subsumed under “overt gender” (cf. Corbett 1991: 44, 62–63, 117–118). That is, NOMINAL FORM CLASSES are established in the present approach by word forms with identical morphological or phonological properties; they represent the counterpart of agreement classes in the realm of morpho(phono)logy. As shown in the important work by Evans (1997) and Evans et al. (1998), nominal form classes (called there “head classes”) can have an intricate relationship to agreement classes well beyond serving potentially as their triggers.

What is called here DERIFLECTION (CLASSES) is the morpho(phono)logical counterpart of genders. They are classes of form paradigms operating over nominal lexemes and established on account of identical formal variation that does not need but often does interact with such features as gender, number, etc. Our newly coined term “deriflection” (a blend of “inflection” and “derivation”) thus refers here in a more narrow sense to relevant morphology or phonology that interacts with gender. In (1) of Swahili the two prefixes on *-toto* ‘child’ establish a specific type of number inflection typical for human nouns, namely *M(W)-/W(A)-*, which is the pairing of a singular and a plural nominal form class exponent. As with genders, deriflections in this context also entail other morpho(phono)logical phenomena to the extent these interact with the relevant nominal system.

In general, agreement class and nominal form class are concepts that relate to a noun as a word form in a concrete morphosyntactic context, while gender and deriflection refer primarily to the more abstract domain of the nominal lexicon in a given language. At the same time, agreement class and gender are both syntactically defined phenomena and thus opposed to nominal form class and deriflection pertaining to the domain of morpho(phono)logy, so that the two concept pairs, although intimately related, are in principle independent from each other. The various interrelations between the four concepts are summarized in Table 1, which also repeats the different notation principles applied for them here.

Corbett’s (1991; 2000; 2006) work has served as the primary reference point for the previous typological analyses of gender and related phenomena. As is to be discussed shortly, however, our framework also departs in some important respects from this author in order to better capture aspects that have subsequently emerged regarding the cross-linguistic diversity in this domain.

Table 1: The four concepts used for analyzing gender

Relates to	Concrete noun in a morpho-syntactic context = word form	Abstract noun in the lexicon = lexeme
Syntax	a. AGREEMENT CLASS (abbreviated as AGR and numbered by Arabic numbers)	b. GENDER (numbered by Roman numbers)
Morpho(phono)logy	c. NOMINAL FORM CLASS (abbreviated as NF)	d. DERIFLECTION

The framework outlined here draws on Güldemann (2000), which dealt with gender systems in Southern African languages of the two non-Khoe families Tuu and Kx’a (both traditionally attributed to a spurious Khoisan lineage). The most important typological contribution of this work is that agreement classes in these languages are often multiply ambiguous regarding their gender and number value, unlike in many European languages, whose analysis has set the stage for the cross-linguistic research on gender and agreement.

AGR	SG		PL
3	<i>ká</i>	V	<i>ká</i>
4	<i>hì</i>	IV	<i>hì</i>
1	<i>ha</i>	II	<i>ha</i>
2		I	<i>sì</i>

Note: agreement classes represented by anaphoric pronouns.

Figure 1: Agreement classes and genders in Jul’hoan (based on Güldemann 2000)

This can be observed in Figure 1, which displays the gender system of the Jul’hoan dialect of Ju, a member of the Kx’a family. The schema shows how the four agreement classes 1–4 pattern across the two number categories singular

(SG) and plural (PL) to yield five genders I–V. The numbering of classes and genders as well as their ordering in the schema are of no concern to the system: the former is an artifact of research history and the latter merely serves to yield a maximally simple representation of the system. The reader is referred to Gülde-mann (2000) for more details, for example, on the semantics of the genders. The only important point for the present discussion is the behavior of the agreement classes, for example, that AGR1 occurs in both number values, singular and plural, as well as in the three genders I–III. The non-sensitivity of an agreement class to number holds in Jul’hoan for AGR1, AGR3, and AGR4. The majority of nouns falling into these classes are not transnumeral but possess different singular and plural forms. Recall from above that a system where the gender marking of nouns only involves one agreement class is as such functionally transparent (albeit typologically rare) in that agreement is here a “non-conflated” *direct* reflex of gender.

The phenomenon that agreement classes are not dedicated to a single gender and/or number is also recurrent outside these Southern African languages, including Niger-Congo. This justifies the strict descriptive and analytical separation of agreement class from any particular value of gender, number etc. This is opposed to Corbett’s (1991) approach, which, moreover, features more analytical concepts than our framework. He distinguishes on the one hand between “controller gender” and “target gender” (see his Section 6.3) and on the other hand between “agreement class” and “consistent agreement pattern” (see his Sections 6.2 and 6.4.5). Our approach, as we argue here, does not need all these notions, because it captures the same data by ascertaining just agreement class (= Corbett’s “consistent agreement pattern”) and gender (= Corbett’s “controller gender”) (our two additional concepts, nominal form class and deriflection, are irrelevant here, because they concern the form of nouns rather than agreement and gender).

Figure 2 takes up Corbett’s (1991: 150–152) example of Romanian adjective agreement, which he uses to illustrate the necessity of his target gender notion. He states about this case that there are “three agreement classes, and there is no reason not to recognize each as a gender [= the lines labeled semantically as masculine, neuter, and feminine]”<sup>5</sup> as well as “two target genders in both singular and plural ... [-Ø, -ă and -i, -e]”. Corbett’s fourth concept, consistent agreement pattern, which we would call agreement class, is not dealt with in his discussion that concerns the exponents of only one agreement context; the notion is,

<sup>5</sup> Although Corbett’s identification of agreement class and gender is surprising, a detailed critical discussion would require a general assessment of his approach, which is beyond the purpose and limits of this paper.

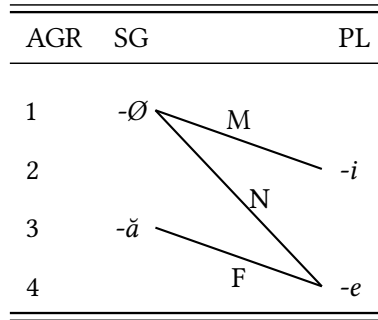


Figure 2: Agreement of adjectives and genders in Romanian (based on Corbett 1991: 152)

however, relevant for a full description, because Romanian has more than one agreement target (see Corbett 1991: 213–214 for further complications in Romanian neuter agreement forms). In any case, Corbett’s problem is that two of the four gender-number markers on adjectives are not dedicated to a single gender,  $-\emptyset$  encoding the singular of both masculine and neuter gender and  $-e$  marking the plural of both neuter and feminine gender; the target gender concept seems to be invoked to solve this problem. However, applying the framework proposed here to the situation in Romanian, we only need to recognize three genders and four agreement classes (representing them here by the four suffixal exponents on adjectives but assuming that other agreement targets do not contradict this picture).

A picture like Figure 2 is nothing special and even in a more extreme case, such as Jul’hoan in Figure 1, it does not require more elaborate analytical machinery. In the Jul’hoan system, comprising five genders across two number values, *three of four* agreement classes are unspecific regarding *both* gender and/or number. As far as we can see, an additional concept like target gender restricted to a specific number category does not furnish any new and useful insight for the description of this and other gender systems. Since the present approach has also been applied with coherent results to a number of other languages with quite different and notoriously intricate gender systems (cf., e.g., Neuhaus 2008 on Krongo of the Kadu family, Güldemann & Maniscalco 2015 on Somali of the Cushitic family), we assume its wider applicability. The rest of the paper attempts to show its usefulness for the languages of Niger-Congo, the world’s largest language family featuring a historically deeply entrenched gender system.

## 2 Niger-Congo gender systems and the philological “noun class” concept

While the noun classification systems in Niger-Congo have long been recognized as instances of grammatical gender, their special structural profile poses particular challenges to a cross-linguistically oriented analysis. To a large extent, this is due to the special morphological characteristics of gender systems in Bantu, the resulting philological tradition of analyzing them, and the considerable research bias within Niger-Congo studies toward this important subgroup (see Güldemann (2018, Chapter 5) for more discussion).

The situation presented in §1 above with example (1) from Swahili is quite typical in Bantu and many other Niger-Congo languages and thus has crucially determined the philological tradition of describing their gender systems as a whole. In particular, it shows a one-to-one relationship between corresponding agreement classes and nominal form classes. As seen in (1b), even the markers can be formally identical: *wa-* (or an allomorph) is the formal exponent in both NF *W(A)-* and all agreement contexts of AGR2. Such a biunique (and often even alliterative) relation between the form of the noun (representing the trigger) and any agreeing element (representing the target) is epitomized by the philological concept of “noun class”. The notion of “noun class” is also behind the philological convention of a single class label by means of Arabic numbers, in opposition to our proposed distinction between agreement class and nominal form class (accordingly, in (1) and subsequent Swahili examples, the nominal form classes are not glossed by Arabic numbers, even in cases of biuniqueness and alliteration).

The conflation of nominal form classes and agreement classes is, as we argue, the reason for a major problem in the analysis of Niger-Congo gender systems. The conceptually overloaded concept of “noun class” may account in many languages for a good portion of the relevant nominal domain, to the extent that the situation is as in (1) of Swahili. However, the concept cannot completely and adequately capture an entire system, because the characteristics implied in it are not universal. Example (1a) with NF *M(W)-* and AGR1 involving *yu-/m-/a-* as its set of exponents has already shown alliteration not to be absolute. More importantly, however, the implied one-to-one relation between agreement classes and nominal form classes also has crucial exceptions so that one type of class is not always predictable from the other, which is shown in the following representative examples.

(2) Swahili (personal knowledge)

- a. *rafiki*      *yu-le*    *m-moja* *a-me-anguka*  
Ø:friend(1) 1-D.DEM 1-one    1-PERF-fall  
‘that one friend has fallen’
- b. *ma-rafiki*    *wa-le*    *wa-wili* *wa-me-anguka*  
MA-friend(2) 2-D.DEM 2-two    2-PERF-fall  
‘those two friends have fallen’

Example (2) shows that Swahili nouns of the human gender, as defined by the pairing AGR1/AGR2, can also appear with other number inflections, here Ø/MA with *rafiki* ‘friend’ (see below for more discussion on prefixless nouns). That is, one agreement class goes with more than one nominal form class.

(3) Swahili (personal knowledge)

- a. *m-ti*              *u-le*      *m-moja* *u-me-anguka*  
M(W)-tree(3) 3-D.DEM 3-one    3-PERF-fall  
‘that one tree has fallen’
- b. *mi-ti*      *i-le*      *mi-wili* *i-me-anguka*  
MI-tree(4) 4-D.DEM 4-two    4-PERF-fall  
‘those two trees have fallen’

Example (3) illustrates that one nominal form class can also be associated with more than one agreement class – the reverse case of the situation illustrated in (2). As shown in (3a), NF *M(W)-* is not exclusively tied to AGR1 in the human gender AGR1/AGR2, as in (1a), but is also relevant for singular forms of lexemes like *-ti* ‘tree’ in AGR3 belonging to the gender AGR3/AGR4. The matching of one nominal form class with more than one agreement class equally holds for NF *MA-* in (2b), because it is also found with plural count nouns of the gender AGR5/AGR6 and with transnumeral nouns for masses and liquids.

To reiterate the point, the philological “noun class” notion inadequately implies the universality of a one-to-one trigger-target mapping, thereby silently conflating the categories of agreement class and nominal form class that are in principle independent. Counterfeiting an ideal system, this concept recurrently decoys scholars into the analytical shortcut illustrated in the following.

Assume a language with gender and nominal deriflection where agreement and nominal form classes display a biunique mapping. Such a situation is represented in Figure 3 (which differs from figures focusing on gender and deriflection

systems such as 1 and 2 above or 4 below). In both domains, the classes are further assumed to map over number such that two apply to singular nouns and one to plural nouns. Such a system would allow one to predict AGR1, AGR2 and AGR3 from NF A, NF B and NF C, respectively, and vice versa – a situation that implies a strong formal assignment of agreement (see Corbett 1991: Chapter 3).

AGR	NF	Number
1	A	SG
2	B	SG
3	C	PL

Figure 3: Full one-to-one mapping of agreement classes and nominal form classes

Figure 4 shows the resulting agreement-based gender system (left side) and the deriflection system based on nominal form classes (right side), which can also be inferred from each other. Here, both show convergence from two singular classes to one plural class. This predictability holds irrespective of whether the exponents in the system of agreement and nominal morphology display alliteration of the type recurrent in Niger-Congo (cf. (1b) from Swahili).

SG	PL	SG	PL
AGR1	AGR3	NF A	NF C
AGR2		NF B	

Figure 4: Gender system (left) vs. deriflection system (right) of the case in Figure 3

In reality, however, an “ideal” trigger-target mapping as in Figure 3 is never universal in a language so that the “noun class” approach harbors the risk of misleading analysis. This can be illustrated by means of a rather well-behaved attested system, like that of Ikaan (Benue-Kwa). Figure 5 shows that there is only a single exception to a complete one-to-one mapping between agreement classes and prefixal nominal form classes, namely NF *O*- that is associated with AGR1 and AGR6. Hence, the system appears to be overall well described in terms of the

canonical unitary concept of “noun class” involving both the forms of nominal prefixes and concords on agreement targets.

With such a neat mapping one may be tempted to proceed according to the discussion revolving around Figures 3 and 4 and infer the agreement-based gender system from the morphological deriflection system based on nominal form classes (or vice versa). Figure 6 shows the two systems side by side for a better comparison. For the record, the two schemas also display a class of transnumeral (TN) nouns marked by circles, which cannot be assigned clearly to a single paired pattern and thus should be recognized as a separate gender. The nature of the various genders and deriflections, including their possible semantics, is largely irrelevant for the present topic and they are therefore not labeled or numbered – a practice also relevant later on, especially in system schemas like those in Figure 6.

The important observation from Figure 6 is that the single exception to a bi-unique class mapping in Figure 5 causes a clear structural divergence between the gender and deriflection systems, as marked by the two thick lines. The difference can be explained in terms of the typology for the mapping of classes across number categories originally proposed by Heine (1982: 196–198) and elaborated by Corbett (1991: 154–158). There are three major types in the order of increasing complexity:

- a. **PARALLEL:** Singular and plural classes only show one-to-one mapping.
- b. **CONVERGENT:** At least two agreement classes in one number converge to one class in the other number.
- c. **CROSSED:** Class convergence exists in both directions.

According to this typology, Ikaan’s real gender system based on agreement is of the convergent type in that the conflation of classes only goes from singular to plural, while its deriflection system based on nominal form classes shows class convergence in both directions and is thus of the more complex crossed type.

In fact, the divergence between gender and deriflection system in Ikaan is almost certainly greater, because the language will have prefixless nouns (e.g., proper names, loans), which are unfortunately not treated in the available sources. These establish an additional nominal form class that does not have a unique counterpart in the agreement system. Since such an additional unmarked  $\emptyset$ -nominal form class can be expected to be virtually universal, this phenomenon alone excludes the one-to-one class mapping and hence the identity of the gender and deriflection system from a general perspective.



## 5 Niger-Congo “noun classes” conflate gender with deriflection

	AGR	NF	Number
6	<i>nò:</i>		SG
1	<i>jò:</i>	<i>O-</i>	SG
2	<i>dà:</i>	<i>A-</i>	TN, PL
3	<i>dò:</i>	<i>U-</i>	SG
4	<i>dè:</i>	<i>I-</i>	SG, PL
5	<i>nè:</i>	<i>E-</i>	SG

Note: agreement classes represented by proximal demonstratives

Figure 5: Mapping of agreement and nominal form classes in Ikaan (based on Borchardt 2011: 75–78)

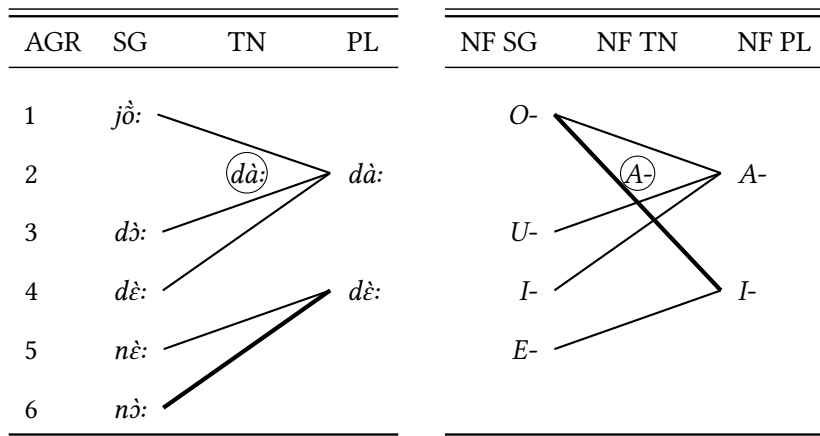


Figure 6: Gender system (left) vs. deriflection system (right) of Ikaan (based on Borchardt 2011: 75–78)

The divergence between gender system and “gender-like” deriflection system holds to an even greater extent in Bantu – the very language group in which the problematic “noun class” concept was developed and from where it assumed its model role for the larger family. This can be illustrated by means of Proto-Bantu for which there exists an elaborate reconstruction. Irrespective of its full historical adequacy, the detailed information of this proto-system allows a good approximation to the original situation regarding (a) the mapping of agreement classes and nominal form classes, (b) the gender system based on agreement classes, and (c) the deriflection system based on nominal form classes.

Excluding an uncertain proto-class \*24, Table 2 presents Meeussen’s (1967: 96–99) full reconstruction of the Proto-Bantu “noun classes”, which, as mentioned,

Table 2: Proto-Bantu “noun classes” (conflating agreement classes and nominal form classes) (based on Meeussen 1967: 96–99)

“Noun class”	Number	AGR	Different agreement targets				NF
			CONC	NUM	SBJ	OBJ	
*2	PL	2	ba-	ba-	ba-	ba-	ba-
*1	SG	1	ju-	u- ?	u-, a-	mu-	mu-
*18	TN	18	mu-	mu-	mu-	mu-	
*3	SG	3	gu-	u- ?	gu-	gu-	
*4	PL	4	gi-	i- ?	gi-	gi-	mi-
*5	SG	5	di-	di-	di-	di-	ĩ-
*6	TN, PL	6	ga-	a- ?	ga-	ga-	ma-
*7	SG	7	ki-	ki-	ki-	ki-	ki-
*8	PL	8	bĩ-	bĩ-	bĩ-	bĩ-	bĩ-
*9	SG	9	ji-	i- ?	ji-	ji-	N-
*10	PL	10	jĩ-	ĩ-	jĩ-	jĩ-	
*11	SG	11	du-	du-	du-	du-	du-
*12	SG	12	ka-	ka-	ka-	ka-	ka-
*13	PL	13	tu-	tu-	tu-	tu-	tu-
*14	TN, SG	14	bu-	bu-	bu-	bu-	bu-
*15	TN, SG	15/17	ku-	ku-	ku-	ku-	ku-
*17	TN						
*16	TN	16	pa-	pa-	pa-	pa-	pa-
*19	SG	19	pĩ-	pĩ-	pĩ-	pĩ-	pĩ-

conflate agreement and noun form. This framework is the outcome of specific developments in Bantu philology, without much consideration of the typological treatment of gender. Hence, it comes as no surprise that it is multiply incompatible with the cross-linguistic approach proposed here.

The divergence between the above Bantu reconstruction and our approach concerns in particular various mismatches between the philological “noun class” inventory in the leftmost column and our analysis that involves the agreement classes in the third column (followed by four columns displaying the exponents of major targets) and the nominal form classes in the rightmost column (we take over the philological class numbering 1–19 for our agreement classes, while nominal form classes are simply referred to by their reconstructed prefix).

The major differences between the Bantu reconstruction and the present analysis, marked in Table 2 by shaded cells, are as follows. First, two nominal form classes, namely those established by the noun prefixes \*mu- and \*N- have a multiple affiliation with agreement classes, the former occurring with nouns of the agreement classes \*1, \*3, and \*18 (cf. the above discussion in connection with (1a) and (3a) from Swahili) and the latter with nouns of the classes \*9 and \*10. Second, two “noun classes” of the Bantu tradition that establish single-class sets of transnumeral nominals should be subsumed under a single noun form and agreement class, because they do not diverge in either nominal prefix or concord. Their difference only concerns the syntactic occurrence of the respective nominal in that “noun class” \*15 comprises infinitives, while “noun class” \*17 is established by the class of general locatives.<sup>6</sup> In general one can conclude that the traditional identification and numbering of “noun classes” in Bantu predominantly target agreement classes. As will be shown in §3, this situation no longer holds for the application of the approach to many other Niger-Congo languages, where the analysis of “noun classes” often, if implicitly, refers to nominal form classes.

Later approaches to Bantu gender systems have introduced yet other conventions that may have enhanced philological comparability but blur cross-linguistic transparency. In particular, Bantuists (and some scholars like Welmers 1973: 166 dealing with other Niger-Congo languages) make an additional “noun class” distinction of \*1 vs. \*1a (and possibly \*2 vs. \*2a). The first class of each pair comprises human nouns with the expected prefix and the latter contains prefixless kinship nouns and proper names. While descriptively adequate, this class differentiation is irrelevant for the inventory of agreement classes but more importantly hides

---

<sup>6</sup>For the record, class \*15 is most likely a grammaticalization from class \*17 via the path locative > purposive > infinitive (cf. Haspelmath 1989).

AGR		NF	Number
X		Ø	SG, PL
*1(a)	u-, a-	*mu-	SG
*3	gu-	X	SG
*18	mu-	X	TN
*2	ba-	*ba-	PL
*4	gi-	*mi-	PL
*15/17	ku-	*ku-	TN, SG
*5	di-	*i-	SG
*6	ga-	*ma-	TN, PL
*14	bu-	*bu-	TN, SG
*7	ki-	*ki-	SG
*8	bi-	*bi-	PL
*9	ji-	*n-	SG
*10	ji-	X	PL
*11	du-	*du-	SG
*12	ka-	*ka-	SG
*13	tu-	*tu-	PL
*16	pa-	*pa-	TN
*19	pi-	*pi-	SG

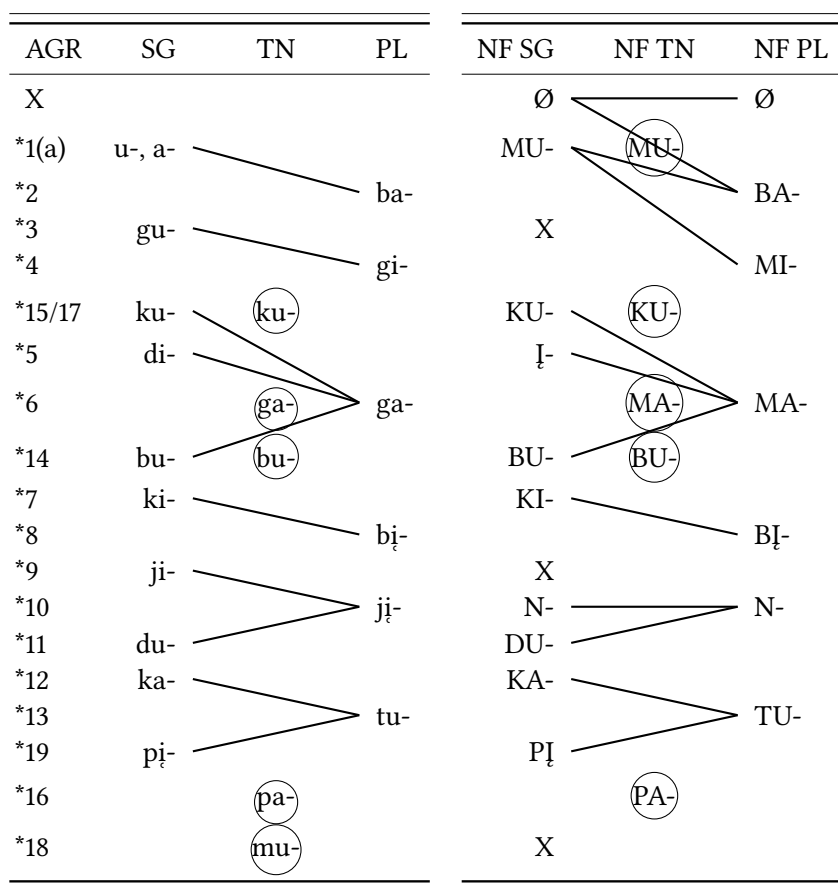
Note: X = no independent class counterpart in the other class type.

Figure 7: Mapping of agreement and nominal form classes in Proto-Bantu

the necessity of taking into account an additional nominal form class Ø that has no unique counterpart in the agreement system (cf. the above discussion in connection with (1a) and (2a) from Swahili).<sup>7</sup>

Figure 7 shows the mapping of agreement and nominal form classes in Proto-Bantu arising from Table 2 (including the additional “noun class” \*1a). Overall, one-to-one trigger-target mapping as well as alliteration are salient but also have important exceptions. The different number of agreement classes and nominal form classes alone, namely 18 vs. 16, implies that the gender system and the derivation system of Proto-Bantu cannot turn out to be completely parallel. In this

<sup>7</sup>See Van de Velde (2006) for an extensive recent discussion of such nouns in Eton and Bantu in general. We do not follow his proposal of considering them as “genderless” nouns, because gender is defined here by agreement and their predominant behavior in this respect clearly assigns them to the human gender.



Note: X = no independent counterpart in the other class type.

Figure 8: Gender system (left) vs. deriflection system (right) of Proto-Bantu

context, the symbol X in this and later schemas stands for the case where no unique counterpart exists for a class in the opposite class type. (The alignment between classes of different type by a horizontal or a sloping line is arbitrary in Figure 7; in the case of historically rooted alliteration, it is useful to connect such etymologically “proper” counterparts by the horizontal line, which will be done in appropriate cases later on.)

A full comparison of the gender and the deriflection systems in Proto-Bantu as reconstructed from the hypothesized “noun class” behavior is shown in Figure 8, which follows the presentation in Figures 4 and 6. In the gender system

on the left side of Figure 8, at least some transnumeral noun groups marked by circles must be analyzed as establishing genders in their own right, because the respective agreement classes cannot be unambiguously associated with a single paired gender, as is the case for AGR6, AGR16, and AGR18 (AGR15/17 and AGR14 are arguably singularia tantum of two paired genders with AGR6 in the plural).

As can be expected, Figure 8 demonstrates considerable differences between the gender and the deriflection system, even more extensively than in Ikaan, despite the still considerable one-to-one alliterative mapping shown in Figure 7. While the gender system with 18 agreement classes is convergent in the above terms and comprises 10 paired genders and at least 3 single-class genders, the deriflection system with 16 nominal form classes is crossed and involves 12 types of morphological number alternations besides 5 types of transnumeral nouns.

Similar or even more dramatic cases of divergence between the gender system and the “gender-like” deriflection system are normal in Niger-Congo, and the problems associated with the traditional “noun class” concept have been recognized in both language-specific and comparative research. The reader is referred to the revealing theoretical and methodological discussion in such studies as Guthrie (1948) for Bantu, and Voorhoeve & de Wolf (1969) and de Wolf (1971) for Benue-Congo. As a consequence, Mieke (forthcoming: 33f) explicitly states that “the marking of nouns and the concord (agreement) systems in their formal and semantic multiplicity should be considered as independent paradigms with regard to their evolution.”

Nevertheless, the philological tradition is so strong that even the only approach known to us that uses the very same analytical concepts as ours yields an analysis that is far from being transparent, namely that by Sterk (1978) for the Nupoid language Gade.

Table 3 betrays hardly any difference to our outline of analytical concepts in Table 1 of §1. The only point is Sterk’s overgeneralization of the singular-plural

Table 3: Sterk’s (1978: 25) concepts for analyzing Gade “noun classes”

	“Prefix” = nominal form class	“Declension” = deriflection	“Class” = agreement class	“Gender”
“form”	+	+	–	–
“concord”	–	–	+	+
“pairing”	–	+	–	+

Note: “...” = Sterk’s (1978) term.

pairing of classes with count nouns in that his last line of the table prescribes the feature “pairing” for “declension” (a.k.a. deriflection) and gender, thus excluding single class patterns with transnumeral nouns.

The real drawback in his description is his complex numbering of “classes”, which aims to cater simultaneously for their morphological shape and their agreement behavior. He writes (ibid.: 27):

We are now faced with the practical problem of how to classify Gade nouns. Noun stems will have to be specified both for declension and for gender, since the one cannot always be predicted from the other. Rather than list noun stems in the lexicon with the double marking however, it is more convenient to devise a system which classifies them unambiguously, both for declension and for gender, with a *single marker*. This will be done by assigning numbers to prefixes, with the proviso: not only will prefixes of differing phonological shape be assigned a different number, but even prefixes of the same shape will be given a different number if the nouns they form part of belong to different [agreement] classes. (emphasis and additions ours).

The single-marker convention proposed by Sterk, which appears to be motivated by the equally conflating “noun class” concept, is the major reason that his presentation falls short of providing a transparent picture of Gade’s nominal system (cf. also Sterk’s (1976) similarly complicated treatment of the Upper Cross language Humono). Our analysis concludes that Gade has a complex deriflection system of more than 30 patterns (albeit many restricted to very few if not a single noun lexeme) based on 13 nominal form classes but a relatively simple system of three productive (and two inoperative) genders based on four regular agreement classes.

replaced  
‘Sterk’s  
single-  
marker  
conven-  
tion’

Comparing the situation in Ikaan, Proto-Bantu, and Gade a potential generalization emerges: in all cases, the agreement-based gender system is simpler (or at least not more complex) than the deriflection system in size and structure – this even if the basic inventory of agreement and nominal form classes shows the opposite picture, as is the case in Proto-Bantu. More data supporting this observation follow in §3 regarding other Niger-Congo languages.

The previous discussion has argued that the Niger-Congo concept and term “noun class” is highly problematic. This is compounded by the fact that the term has come to bear different meanings in Niger-Congo studies, depending on diverse language-specific situations. Thus, in languages that lost (most of) the inherited agreement, “noun class” may just refer to nominal form classes, as in

some Gur languages, for example, Moore (Canu 1976), or in the Idomoid language Igede (Abiodun 1989) (see also Good 2012: §4.2). In a parallel fashion, in the apparently rarer case of the loss of transparent noun affixes with retention of agreement, the term “noun class” tends to mean merely agreement class, as is the case to varying degrees in Wolof from Atlantic (Babou & Loporcaro 2016) and Mundabli from Bantoid (Voll 2017) (see also Good 2012: §4.3). Finally, the discussion in §3.2 below about Akan shows that some authors even use “noun class” for deriflection (class). From a global typological perspective, yet another complication arises from the terminological tradition in other geographical areas: in Caucasian and partly Australian languages, the term “noun class” refers to gender. The same usage has been proposed by Aikhenvald (2000) for typological investigation in general, the term “gender” being restricted to sex-based systems. We consider this proposal to be unfortunate because it not only diverges from Corbett’s (1991) earlier and widely accepted terminology but also disregards the fact that in Niger-Congo, the largest language family on the globe where “noun class” plays a central role, it does conventionally not refer to gender (pace the statement in some relevant studies, e.g., Kilarski 2013: 1). In view of the multiple ambiguity of the term “noun class”, covering in fact all the four analytical concepts outlined in §1, we do not use the term in any other meaning than the original philological one in Niger-Congo and employ it in quotation marks for the sake of clarity.

### 3 Examples for the treatment of individual Niger-Congo groups

#### 3.1 Introduction

As was said above, the approach to Niger-Congo gender and deriflection systems in terms of “noun classes” has been and still is the rule. In the following we show that as a result analyses of individual languages and attempted reconstructions of language groups<sup>8</sup> often deal predominantly or exclusively with the

---

<sup>8</sup>Until now, (partial) reconstructions of gender and deriflection systems exist for relatively few of the numerous Niger-Congo groups. In addition to Bantu, we are aware of those for Gur (Manessy 1967, 1975; Mieke et al. 2012), Ghana-Togo-Mountain (Heine 1968, see §3.3), Benue-Congo (de Wolf 1971), Mbaic (Bokula 1971, Pasch 1986), Atlantic (Doneux 1975), Non-Bantu Bantoid (Hyman 1980), Edoid (Elugbe 1983), Lower Cross (Connell 1987), and Guang (Manessy 1987, Snider 1988, see §3.4). In addition, comparative treatments exist on groups that are uncertain members of Niger-Congo (see Güldemann 2018) but have typologically similar nominal systems such as Heibanic (Schadeberg 1981a), Talodic (Schadeberg 1981b), and Kru (Marchese 1988).



system of number inflection rather than gender. We demonstrate and elucidate this mistaken approach with data from Akan (§3.2), Guang (§3.3), and Ghana-Togo-Mountain (§3.4). These geographically close but structurally sufficiently diverse Niger-Congo groups in West Africa that are commonly subsumed under the ambiguous concept of Kwa (see Güldemann 2018 for more discussion on the problematic genealogical classification) represent a convenience choice. The discussion would hardly differ by using other Niger-Congo groups and our approach has indeed been applied with the same results to other relevant languages, for example, Kisi, Wolof, Fula, and Laala from Atlantic, Miyobe from Gur, C’lela and Gade from Benue-Kwa, and Mbane from Ubangi.

We will proceed in our analysis according to the framework outlined in §1. For each language (or proto-language), we first present the agreement class system in the form of a table. This table represents each class by means of exponents in the most important agreement targets, records its behavior regarding number, and, if applicable, gives the default nominal form class. We number the language-specific classes by Arabic numbers either according to the source or our own arbitrary choice; these numbers are preceded by an acronym of the language in order to avoid any facile association with the comparative Bantu-Niger-Congo system. The gender system is established on the basis of the attested mapping of these agreement classes over the relevant number categories and presented in the form of a figure. Agreement classes are represented by one maximally distinct agreement target, similar to previous schemas; genders only receive a label in systems with few distinctions and reasonable clear semantics. Salient sets of transnumeral nouns are marked as usual by circles in the structural schemas; those that cannot be assigned to a paired-class gender in a straightforward way would establish separate single-class genders. Doubtful genders, including “in-quate” ones in terms of Corbett (1991: 170–175), that is, agreement-based sets of nouns whose small size is arguably insufficient to merit incorporation into the grammatical gender system, may be marked by broken lines or circles. This practice is at best approximate, as the available data are insufficient; notably because they usually do not give a full picture about lexical frequencies. In general, the following overviews of gender (and deriflection) patterns are “structural” systems that may have to be changed with more comprehensive information about the entire nominal lexicon of a language.

The description of the agreement and gender systems is followed by the investigation of nominal form classes and the resulting deriflection system. Nominal form classes, which are represented by an abstract thematic element in capital letters, are also given in a table that includes their number behavior and representative sample nouns. As far as possible, we take the Ø-marked class (e.g.,

loans, personal names, kinship terms) into account. The deriflection system is represented in a parallel fashion to the gender system.

Finally, in order to elucidate the relationship between gender and deriflection system, we discuss the discernible correspondences and mismatches between agreement and nominal form classes. These are schematized in figures similar to those given above. In doing so, we try to reflect, if appropriate, the original (al-literative) match between agreement and nominal form class, which is assumed to originate in an older Niger-Congo state and whose best proxy at the present is still the relatively coherent Proto-Bantu system.

The following discussion involves at several places an assessment of Niger-Congo systems regarding a notion of complexity that differs from that focussed on in §2, which was concerned with systemic organization. In line with Di Garbo's (2014: 41, 179) first principle of absolute complexity, the characterization here ascertains a system's number of genders (and deriflections). Our evaluation is done against the background of the widely assumed Proto-Niger-Congo state, which, when modeled on Bantu, would have involved around ten or even more distinctions in both domains, as well as Corbett's (2013) typological approach, which assigns the label "complex" to gender systems with five or more distinctions. That is, we consider a Niger-Congo system as reduced (or no longer as complex), if its inventory has been decreased to a value lower than Corbett's typological threshold for his highest degree of complexity. Note the partly misleading bias toward this typological standard, because a system with five genders like in Logba (Ghana-Togo-Mountain) is certainly reduced vis-à-vis the proto-state but still counts here as complex.

### **3.2 Akan**

Akan is the first linguistic entity to be discussed. It is a large language complex that is the core of a group of closely related languages called Akanic, which in turn is classified under the Potou-Akanic family (Stewart 2002). Akan's most important dialects in Ghana are Akuapem, Fante and Asante (Dolphyne & Dakubu 1988: 57).

The evaluation of the synchronic nominal system of Akan undertaken by various authors differs considerably, and none transparently captures the full picture of a system with complex number inflection and, in some dialects, a simple animacy-based gender system. We argue that this is due to a large extent to the problematic philological Niger-Congo tradition outlined in §2.

Earlier authors like Christaller (1875), Dolphyne & Dakubu (1988), etc. recognize nominal prefixes in Akan but do not relate these to a nominal system of

the Niger-Congo type, thus failing to identify any possible grammatical aspect of “noun classes”. Following Welmers’ (1971: 4–5) short notes, Osam (1993) is possibly the first author who analyzes the nominal prefixes as vestiges of a formerly complex “noun class” system. Equally important is that the author also discusses agreement phenomena that are arguably remnants of the inherited Niger-Congo gender system. Given the focus of this paper, these need to be outlined in more detail.

For one thing, there is number agreement between nouns and a sub-group of attributive adjectives in that the latter receive a prefix in the plural. The nasal prefixes on both the trigger and the target in example (4b) suggest that there is correspondence in gender and number between the pluralized noun and the modifying adjective.

(4) Akan (Osam 1993: 98, 87)

- a. *a-bofra kakramba*  
A-child small  
‘small child’
- b. *m-bofra n-kakramba*  
N-child PL-small  
‘small children’

The author’s explanations and additional examples as that in (5) make it clear, however, that formal prefix identity as in (4b) is coincidental. Although this is not stated explicitly, the available data suggest that plural marking on adjectives is lexicalized and thus independent of the noun, so that synchronically this phenomenon does not entail gender.

(5) Akan (Osam 1993: 98)

- a. *a-kyen n-kakramba*  
A-drum PL-small  
‘small drums’
- b. *n-tar e-tuntum*  
N-dress PL-black  
‘black dresses’

However, some Akan dialects like Fante and Bron also display verbal subject cross-reference in which the agreement with the relevant nominal referent operates according to the feature of animacy, as shown in (6) for singular number and systematized in the full picture of Table 4.

(6) Akan (Osam 1993: 93)

- a. *ɔ-bɛ-yera*  
1-FUT-be.lost  
‘s/he will be lost’
- b. *ɛ-bɛ-yera*  
3-FUT-be.lost  
‘it will be lost’

Table 4: Agreement system of some Akan dialects (based on Osam 1993)

AGR	Number	Verb prefix	Semantics
AK1	SG	ɔ-, o- = O-	animate
AK2	PL	wɔ-, wɔ- = wO-	animate
AK3	SG, PL	ɛ-, e- = E-	inanimate

Note: multiple forms due to vowel harmony.

Despite the data presented, Osam’s (1993: 99–100, 102) major conclusions are that modern Akan “does not have a functioning noun class system” nor “a concordial system”, whereby he presumably refers to such elaborate and productive ones as in Bantu and similar Niger-Congo groups. From a typological perspective, however, Akan dialects like Fante and Bron must be analyzed as having a gender system that is structurally of the parallel type and semantically driven by a distinction of animate vs. inanimate nouns, as shown in Figure 9.

AGR	SG		PL
AK1	O-	AN	
AK2			wO-
AK3	E-	IAN	E-

Figure 9: Gender system of some Akan dialects (based on Osam 1993)

Bodomo & Marfo (2006) is another study dealing with the nominal system of Akan. These authors explicitly contradict one of Osam’s conclusions in identifying a functional “noun class system” on account of nominal affixation, which

not only involves prefixes but also suffixes. As just another token of the theoretical and terminological confusion in Niger-Congo studies, “noun classes” in their terms are sets of nouns showing the same singular/plural affix pairing, that is, classes of number inflection or deriflection in the above, and for that matter common typological, approach. The authors describe a complex system of 9 “noun classes” a.k.a. deriflections, which partly involve class pairs and subclasses. This is schematized in Figure 10 (restricted to prefixes) and exemplified fully in Table 5.

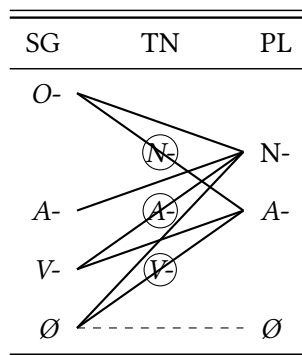


Figure 10: Deriflection system of Akan (based on Bodomo & Marfo 2006)

As can be seen in Table 5, some of the authors’ “noun classes” a.k.a. deriflections, namely 5, 6 and 7, which all relate to various types of human nouns, involve suffixes in addition to prefixes. Except for the pattern 5b, these suffixes do not create deriflection types that do not already exist on account of the 5 prefix-based nominal form classes. For this reason we only integrate the new Ø/Ø prefix pattern (see the broken line) in our analysis of the deriflection system in Figure 10. This system involves 8 patterns for count nouns and three for transnumeral nouns. From a structural perspective, it is a complex crossed system because all types of singular noun forms except for the A-class combine with the two productive plural form classes N- and A-.

As discussed above, only some varieties of Akan have a parallel system of two genders. Here, the inventory of three agreement classes is so reduced that any correspondence between these and the numerous nominal form classes can only be limited. In fact, the only clear match in both form and meaning exists between AK1 with the exponent O- and NF O-; both mark (predominantly) animate singular nouns. Obviously, this situation diverges considerably from the picture involving “noun classes” of Bantu-type languages, which involve both agreement and morphological form.

Table 5: Deriflection system of Akan (based on Bodomo &amp; Marfo 2006: 214–217)

“Noun Class” a.k.a. deriflection	Example(s) Meaning	SG	TN	PL
1: V-/N-	–			
a: O-/N-	‘female’	ɔ̀-bàà		m-máá
b: A-/N-	‘cloth’	à-tààdé		n-tààdé
c: (V)-/N-	‘time’	è-bré		m-mré
2: Ø-/N-	–			
	‘mountain’	béɔ́		m-méɔ́
3: V-/A-	–			
a: O-/A-	‘elephant’	ɔ̀-sónɔ́		à-sónɔ́
b: (V)-/A-	‘house’	è-fíé		à-fíe
4: Ø-/A-	–			
	‘veranda’	bámá		à-bámá
5: (V-)/(A-)_-nɔ́m	<b>Kinship</b>			
a: V-/A-_-nɔ́m	‘father’	à-gyá		à-gyá-nɔ́m
	‘wife’	ɔ̀-yíí		à-yíí-nɔ́m
b: Ø-/Ø-_-nɔ́m	‘aunt’	sèwáá		sèwáá-nɔ́m
6: (O)-_-ni/A-_-fɔ́ɔ	<b>Identity/occupation</b>			
a: O-_-ni/A-_-fɔ́ɔ	‘Christian’	ò-kristò-ní		à-kristò-fúɔ́
b: Ø-_-ni/A-_-fɔ́ɔ	‘teacher’	tíkyà-ní		a-tíkyà-fúɔ́
7: (O)-_(-ni)/N-_-fɔ́ɔ	<b>Identity</b>			
a: O-_-ni/N-_-fɔ́ɔ	‘Muslim’	ò-krè̀mò-ni		n-krè̀mò-fúɔ́
b: O-_-Ø/N-_-fɔ́ɔ	‘ghost’	ɔ̀-sámáń		n-sámàn-fúɔ́
8: A-	<b>Deverbal derivation</b>			
	‘farming’		à-dɔ́	
9: N~V-	<b>Mass</b>			
a: N-	‘water’		n-su	
b: V-	‘fire’		è-gyá	

In summary, the Niger-Congo tradition clearly fails to capture the structures encountered in Akan. Its conceptual framework has even misled descriptive linguists, although the picture as such is not hard to understand as involving a complex, semantically sensitive deriflection system and in some dialects a far simpler agreement-based gender system steered by animacy. As for Osam (1993), he fails to clearly identify both phenomena in spite of providing most of the relevant empirical data. Bodomo & Marfo (2006: 206), in turn, state that “[a]n overview of ... nominal morphology shows that the most appropriate criterion that can be used to set up noun classes is number – i.e. singular and plural – categorization”, while “concord marking ... is not a very sufficient criterion”. They thus acknowledge that mainstream Akan has a system of overt noun classification by means of nominal morphology but fail to observe explicitly that this type of nominal categorization is crucially different from gender in general and the original Niger-Congo system in particular (this apart from not dealing with the animacy-based gender system in some dialects).

### 3.3 Guang

#### 3.3.1 Introduction

The second language group we deal with is the Guang family, which like Akanic belongs to the larger Potou-Akanic lineage within Benue-Kwa. Guang languages are known for their elaborate nominal prefix system but are said to show little in the way of agreement.

In all the Guang languages, singular and plural of nouns is [sic] indicated by prefixes. None exhibit concord systems, such as are found in many of the Central Togo languages [= Ghana-Togo-Mountain, cf. §3.4]. There is, however, at least a trace of number agreement between the noun and some types of adjectives in South Guang, Gichode, Krachi, and some dialects of Nchumburu ... (Dolphyne & Dakubu 1988: 82)

Most attempts to define Guang “class” systems are thus restricted to nominal form classes and disregard concord (and the potentially resulting genders). Our ongoing research aimed at a typologically informed survey of the Guang family reveals that the picture, summarized in Table 6, is in fact far more diverse.

Table 6 shows that gender agreement is indeed strongly reduced in several Guang languages, largely to an animacy differentiation illustrated in Figure 11 with the case of Gonja, which is parallel to the situation in the relevant Akan dialects treated in §3.2. However, several languages still possess quite complex gender systems, for example, Chumburung, which we illustrate in §3.3.2.

Table 6: Overview over gender systems in Guang

Languages	Gender agreement	Number inflection
Chumburung, Foodo, Gichode, Ginyanga, Nawuri	complex	complex
Awutu, Dwang, Gonja, Gua, Krache, Larteh, Nkami, Nkonya	reduced	complex
Cherepon, Dampo, Nterato, Kplang, Nchumbulu, Tchumbuli	insufficient information	insufficient information

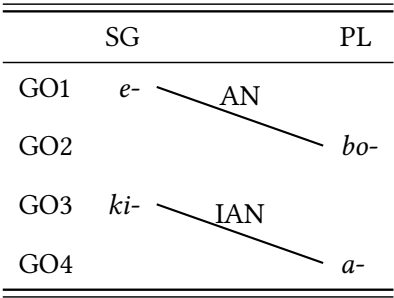


Figure 11: Gender system of Gonja (based on Painter 1970)

3.3.2 Chumburung

Chumburung, according to the description by Hansford (1990: 266ff), is a Guang language with a more canonical nominal system. Its agreement system concerns both the noun phrase in the form of quantifier agreement, as in (7), and a variety of other morpho-syntactic contexts with anaphoric pronominal agreement, for example, the conjoined noun phrase in (8). Other targets of the second type of concord are pronominal forms for ‘certain’, ‘one of’, ‘each, any’, ‘which’ and demonstratives (Hansford 1990: 184); when these are used as modifiers within a noun phrase, they do not agree with their head. A similar situation holds for verbal subject and object cross-reference and relative clauses, as in (9) (Hansford 1990: 450). The full system of seven agreement classes is provided in Table 7.



- (7) Chumburung (Hansford 1990: 270, 201)  
 a. *ǎ-wàǎgyà didáá á-nyó mǎ*  
 A-cloth(6) old 6-two DEM  
 ‘these two old cloths’  
 b. *í-wóri í-nyó í-nyó*  
 I-book(4) 4-two 4-two  
 ‘pairs of two books’ (distributive)
- (8) Chumburung (Hansford 1990: 266)  
*wàǎgyà gyígyíí nà ó-pípéé*  
 Ø:cloth(1) black and 1-red  
 ‘a black and red cloth [lit.: a black cloth and a red one]’
- (9) Chumburung (Hansford 1990: 267, 451)  
*kí-bígyá ní kǐ í gyí sí ó*  
 KV-side(3) REL 3 IPFV eat on REL  
 ‘... the side that will win’

Table 7: Agreement class system of Chumburung (based on Hansford 1990)

AGR	Number	SBJ	OBJ	Pronominal	NF default
CH1	TN, SG	<i>ǎ-/o-</i>	<i>`-</i>	<i>ǎ-/o-</i>	–
CH2	TN, PL	<i>bǔ-/ba-</i>	<i>bá-</i>	<i>bǔ-/ba-</i>	–
CH3	TN, SG	<i>kV-</i>	<i>kí-</i>	<i>kǔ-/kǐ-</i>	<i>KV-</i>
CH4	TN, PL	<i>i-/ǐ-</i>	<i>í-</i>	<i>ǐ-</i>	<i>I-</i>
CH5	TN, SG	<i>ka-</i>	<i>ká-</i>	<i>ka-</i>	<i>KA-</i>
CH6	TN, PL	<i>a-</i>	<i>á-</i>	<i>a-</i>	<i>A-</i>
CH7	TN, PL	<i>N-</i>	<i>n̄-</i>	<i>ŋ-/m-</i>	<i>N-</i>

While Hansford does not give a schematic overview of the gender system, his description of the mapping of agreement classes over number categories allows one to establish the system in Figure 12 with six paired and at least four single-class genders.

	SG	TN	PL
CH2		(bV-)	bV-
CH1	O-	(O-)	
CH4		(I-)	I-
CH3	kV-	(kV-)	
CH6		(a-)	a-
CH5	ka-	(ka-)	
CH7		(N-)	N-

Figure 12: Gender system of Chumburung (based on Hansford 1990)

When compared to the widely assumed Niger-Congo proto-type, this complex crossed system is in several respects remarkable, which is largely due to the nature of agreement classes in Chumburung. For one thing, all agreement classes occur with transnumeral nouns, so that at least some are not dedicated to a single number feature. For CH2, CH5, and CH7, one may avoid positing separate single-class genders by arguing that these nouns represent special transnumeral cases, namely singulalia tantum or pluralia tantum that can be associated uniquely with particular paired genders, namely CH1/CH2 and CH5/CH7. However, this solution is not possible for similar nouns in the remaining four agreement classes, because it would be an ad-hoc decision at this stage to assign these nouns to one of the two or even three paired genders the relevant class partakes in. The last fact is another non-canonical finding in the present philological context, namely that only the three aforementioned classes, CH2, CH5, and CH7, have a unique counterpart in their opposite number feature and are thus dedicated to a paired gender. Overall, Chumburung agreement classes only poorly meet the Niger-Congo expectation that “noun classes” only have one number and one gender value.

The system of seven nominal form classes described for Chumburung, including the group of prefixless nouns, are exemplified in Table 8, while Figure 13 displays their mapping over number categories in the deriflection system.

The deriflection system, presented by Hansford with example nouns, comprises 7 types of singular-plural pairings, and all nominal form classes also occur with transnumeral nouns. Although this crossed system is overall similar in

Table 8: Nominal form class system of Chumburung

NF	Form	Examples
Ø	–	SG <i>dáá</i> ‘elder brother’, <i>bóri</i> ‘voice’ TN <i>gyàbwí</i> ‘honey’, <i>sáŋ</i> ‘time’
O-	o-/ɔ-	SG <i>ó-wúrè</i> ‘chief’, <i>ɔ-dɔ́</i> ‘fishing net’ TN <i>ɔ-tòrí</i> ‘morning star’
I-	i-/ɪ-	TN <i>í-bírísí</i> ‘evil spirit(s)’ PL <i>í-bóri</i> ‘voices’, <i>ì-dɔ́</i> ‘fishing net’, <i>ì-síbɔ́</i> ‘ears’, <i>í-bá</i> ‘coming (PL)’
KV-	ki-/kɪ-/ ku-/kʊ-	SG <i>kì-yéé</i> ‘meat’, <i>kì-síbɔ́</i> ‘ear’, <i>kí-bá</i> ‘coming’ TN <i>kì-tìrì</i> ‘poverty’
A-	a-	TN <i>à-bání</i> ‘government’ PL <i>á-dáá</i> ‘elder brothers’, <i>á-wúrè</i> ‘chiefs’, <i>à-yéé</i> ‘meats’
KA-	ka-	SG <i>ká-mé</i> ‘stomach’ TN <i>ká-nyíté</i> ‘patience’, <i>ká-kyínà</i> ‘life’
N-	n-/m-/ ŋ-/ɲ-	TN <i>m-bɔ́gyà</i> ‘blood’, <i>m-bèráá</i> ‘law’ PL <i>m-mé</i> ‘stomachs’

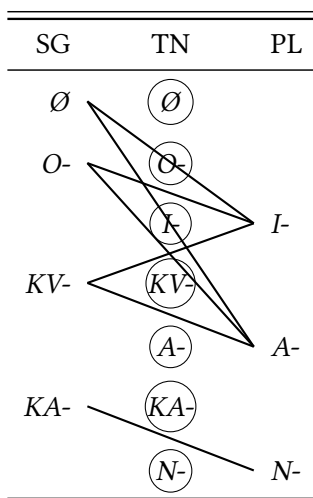


Figure 13: Deriflection system of Chumburung (based on Hansford 1990: 156–161)

structure and size to the gender system in Figure 12 with 6 paired and 4 single-class patterns, it is more complex than the latter on account of having 7 paired deriflections.

	AGR		NF	Number
	X		Ø	TN, SG
CH1	O-		O-	TN, SG
CH3	kV-		kV-	TN, SG
CH5	ka-		kA-	TN, SG
CH4	I-		I-	TN, PL
CH6	a-		A-	TN, PL
CH7	N-		N-	TN, PL
CH2	bV-		X	PL

Note: X = no independent class counterpart in the other class type.

Figure 14: Mapping of agreement and nominal form classes in Chum-burung (based on Hansford 1990: 156–161)

The concrete differences between the systems of genders and deriflections are due to a number of mismatches between agreement and nominal form classes, as shown in Figure 14. These exist in spite of the still considerable formal correspondence between the two sets that is expected from the inherited one-to-one alliterative mapping. A predictable mismatch is the existence of the Ø-nominal form class that has no independent match in the agreement system. Another difference arises from the loss of the reconstructable nominal form class counterpart of CH2; the relevant nouns are found today in two other nominal form classes in A- (a potential reflex of the expected prefix \*ba- through loss of the initial consonant) and N-. Both points are related to another important phenomenon also found in other Guang languages; namely that the semantic criterion of animacy overrides the inherited, more elaborate formal gender assignment. That is, all human nouns irrespective of their form class prompt agreement according to singular CH1 and plural CH2 (the nominal form class in I- is the only one without human nouns). The power of this semantic criterion can also be seen when analyzing the agreement triggered by proper nouns: all singulars agree according to CH1; all plurals referring to humans, personified animals and supernatural beings belong to CH2 while the rest follows CH4 or CH6 (Hansford 1990: 166).

### 3.3.3 Proto-Guang

The “noun class” system of the Guang family has been subject to historical-comparative reconstruction independently but roughly at the same time by Manessy (1987) and Snider (1988). We discuss their results in the following before the background and in accordance with the presentation of our Chumburung analysis in the Figures 12 and 13.

As already suggested by Manessy’s term “système classificatoire” (instead of “gender system”), this author takes both nominal form classes and agreement in the pronominal system of some languages into account, although the latter was at his time only available for two languages, namely Nkonya (Westermann 1922, Reineke 1966) and Gonja (Painter 1970). For all other languages, he merely had access to wordlists that only rarely contained information on agreement. A yet greater problem of his analysis is that he follows the philological approach in explicitly (ibid.: 42) conflating noun form and agreement classes into a single Guang reconstruction, given in the left schema of Figure 15.

Snider (1988) deduced the “noun class” system of Proto-Guang by looking at the noun prefixes of nine of the 18 attested family members without mentioning at all possible agreement forms. He observed a major difference between Northern and Southern Guang, the former being richer in nominal form classes, and concluded (ibid.: 138):

... that proto-Guang had a system at least as complex as the most complex present day Guang languages and that the southern Guang languages represent a collapsing of classes.

The system he established for Proto-Guang is displayed in the middle of Figure 15; we have added the three single-class patterns mentioned by him when discussing the individual nominal form classes.

We briefly show in the following that both Proto-Guang systems in Figure 15 are biased toward the situation in other West African class languages and/or the authors’ assumptions about Proto-Niger-Congo. Moreover, nominal form classes are the primary source for the analysis, even though agreement classes are taken into account to some extent. This bias and the conflation of all data into a single “noun class” system causes serious errors in their reconstruction results, so that they not only differ from each other but also both fail to yield a likely approximation to either the gender or the deriflection system of Proto-Guang. The last point is evident from an inspection of the gender system in such modern languages as Chumburung (repeated from §3.3.2 on the right side of Figure 15).

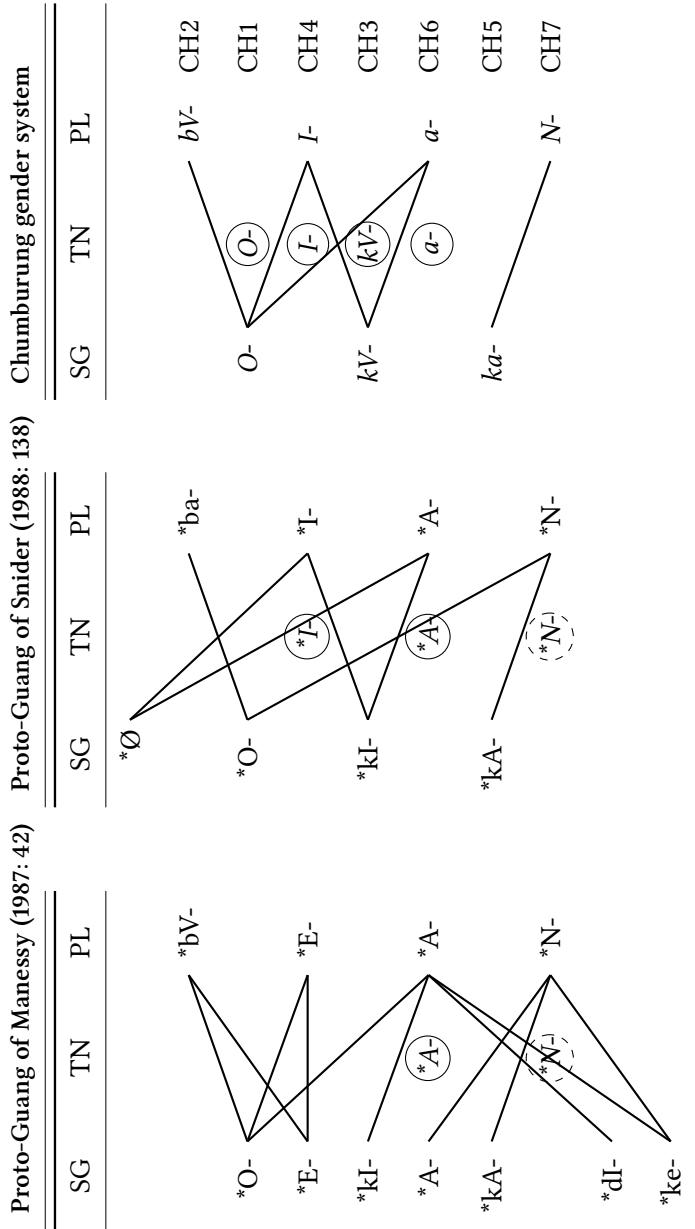


Figure 15: Noun classification systems of Proto-Guang and Chumburung

The following can be observed regarding the (non)overlap between the two proto-systems. Manessy and Snider only agree on the three class pairs \*kI-/A-, \*ka-/N-, and \*O-/bV-, all of which are also attested as genders in modern Chumburung. Both Manessy (1987: 27) and Snider (1988: 141) reconstruct a plural prefix \*bV- or \*ba-, although they observe its exceptional status in that it only occurs as such in Gonja; they claim it to belong to the proto-language because of its wide distribution in Niger-Congo as well as its attestation as an agreement form for third-person plural (animate) in a range of Guang languages.

Snider reconstructs a Ø-class but merely as part of the number inflection patterns \*Ø/I- and \*Ø/A- without noting that these reflect agreement-based genders that in the singular involve the old Niger-Congo class \*1, as can be observed in modern Chumburung (his additional nominal prefix pairing \*O-/N- is so far not attested as involving a separate gender). Although Manessy (1987) appears to capture well the behavior of the old Niger-Congo class \*1, he does not posit a Ø-class for nouns. According to him, most prefixless nouns in one language show a *kV*-prefix in another language, concluding that in the proto-language such nouns did not form a “noun class” (Manessy 1987: 20); in our view this seems to be adequate with respect to agreement while not being the case for noun forms.

Another major divergence between the two reconstructions concerns all forms in *kV*-. Snider (1988: 147–148) reconstructs the prefixes \*kA- and \*kI- (representing *ki*-, *kɪ*-, *ku*-, and *kʊ*-). Manessy (1987: 12) additionally posits \*ke- (representing *ke*-, *kɛ*-, *ko*-, and *kɔ*-), assumed by Snider to be due to phonetically inaccurate data. All Guang languages only have a binary distinction of *kV*-forms in the agreement system but, due to the complexity of the vowel phonology, dispose of a wider range of relevant forms on nouns. Thus, Manessy’s two class pairs based on a third \*ke- do not seem to be warranted, because they are only attested in Gichode (and probably Ginyanga) as genders and deriflections in opposition to a *gI*-class, so that putative \*ke- may merely be a reflex of \*kA-.

Manessy’s Proto-Guang reconstruction is problematic in several other respects. His pair \*E-/bV- only exists as a gender and deriflection in Gonja (see Figure 11). He also posits a singular prefix \*dI- (paired with plural \*A-), although it is only attested in such a gender in Foodo (which was not part of Snider’s language sample). Manessy includes \*dI- for Proto-Guang, because there are nouns with a purported *IV*-prefix in some other Guang languages and the prefix is “fort commune dans les langues à classes d’Afrique occidentale et que pour cette raison nous tenons pour ancienne [very common in the class languages of West Africa and for that reason we consider to be old]” (Manessy 1987: 41). His reconstructions \*E-/E- and \*A-/N- are not attested genders in any language and are also

questionable as reconstructable deriflections. Finally, he fails to identify the pairing \*kI-/E-.

A general conclusion about Manessy's and Snider's historical-comparative work on Guang is that their philological approach generates reconstructions that reflect the agreement and resulting gender system inadequately. In particular, their focus on nominal form classes seems to result in proto-systems that are overly complex for the domain of genders.

### 3.4 Ghana-Togo-Mountain

#### 3.4.1 Introduction

The Ghana-Togo-Mountain languages (formerly known as Togo Remnant) are spoken in Ghana, Togo and Benin. Besides the relevant Guang languages, they are well known within Kwa for class systems that retain both rich agreement and noun prefix patterns. Historical comparisons across these languages are complicated by their unresolved genealogical classification in that they are viewed either as a single lineage according to the traditional view or as forming at least two families according to more recent research (cf. Blench 2009 for a relevant discussion). Table 9 shows the subclassification of the languages after Hammarström et al. (2018) and the profile of their noun categorization systems according to Güldemann & Fiedler (2016).

Table 9: Inventory, classification and noun categorization profile of Ghana-Togo-Mountain languages

	Language(s)	Gender agreement	Number inflection
Na-Togo	Anii, Adele, Lelemi, Siwu, Sekpele, Selee, Logba	complex	complex
	Boro (†)	no information	no information
Ka-Togo	Avatime, Nyangbo, Tafi, Tuwuli, Akebu	complex	complex
	Igo, Animere	reduced	complex
	Ikposo	absent	absent



As with Guang in §3.3, we will first present the synchronic gender system of one modern Ghana-Togo-Mountain language before turning to historical approaches to the entire group.

### 3.4.2 Lelemi

We have chosen the Na-Togo language Lelemi (as described by Allan 1973 with a focus on the Baglo variety) as an example, because it possesses a complex gender system and it has also been included in the typological gender survey by Corbett (1991).

Lelemi nouns prompt agreement on a variety of targets such as determiners, as in (11), ordinal numerals, the cardinal numeral ‘one’, participles, as in (10), and relative pronouns, as well as anaphoric subject cross-reference, as in (11). As opposed to Heine (1968: 115), Allan’s data do not provide evidence for adjectival agreement.

- (10) Lelemi (Allan 1973: 178)

*kɔ̀-làkpi kɔ̀-dun-di*  
 KO-snake(6) 6-kill-PART  
 ‘a killed snake’

- (11) Lelemi (Allan 1973: 240–241)

<i>ɔ̀-nàná</i>	<i>ɔ̀-mè</i>	<i>ɔ̀-dìà</i>	‘this man
<i>bà-nàná</i>	<i>bá-mè</i>	<i>bà-dìà</i>	‘these men
<i>lɛ-tɔ</i>	<i>lé-mè</i>	<i>lè-dìà</i>	‘these houses
<i>a-nimì</i>	<i>á-mè</i>	<i>à-dìà</i>	‘this rice
<i>kɔ̀-di</i>	<i>kɔ̀-mè</i>	<i>kɔ̀-dìà</i>	‘this cloth
<i>ke-mo</i>	<i>ká-mè</i>	<i>kà-dìà</i>	‘this farm
<i>n-tɛ</i>	<i>bɔ̀-mè</i>	<i>bɔ̀-dìà</i>	‘this palm wine
NF-x	AGR-this	AGR-be.good ...	is/are good’

Table 10 summarizes the agreement system of Lelemi. Different from Allan (1973) we posit one more agreement class, LE4, for plural nouns with a prefix *LE-*, because these display a distinct set of concord exponents, which is intermediate between that of LE3 and LE5 (cf. bold-faced elements in the table).

The gender system is not given by Allan (1973) but can be deduced from the relevant behavior of agreement classes. Figure 16 shows that it comprises 9 paired and 7 single-class patterns.

Table 10: Agreement class system of Lelemi (based on Allan 1973)

AGR	Number	DEM/REL SBJ/PART*	POSS	OBJ	PRO	NF default
LE1	TN, SG	<i>ɔ-/u-</i>	<i>ɲwa</i>	<i>ɲ</i>	<i>àɲu</i>	–
LE2	PL	<i>ba-/be-</i>	<i>Bana</i>	<i>mà</i>	<i>àma</i>	–
LE3	SG	<i>le-/li-</i>	<i>anya</i>	<i>nì</i>	<i>àni</i>	<i>LE-</i>
LE4	TN, PL	<i>le-/li-</i>	<i>anya</i>	<i>nyà</i>	<i>ànya</i>	<i>LE-</i>
LE5	TN, PL	<i>a-/e-</i>	<i>ana</i>	<i>nyà</i>	<i>ànya</i>	<i>A-</i>
LE6	all	<i>kɔ-/ku-</i>	<i>kuna</i>	<i>kù</i>	<i>àku</i>	<i>KO-</i>
LE7	TN, SG	<i>ka-/ke-</i>	<i>kana</i>	<i>kà</i>	<i>àka</i>	<i>KA-</i>
LE8	TN, PL	<i>bɔ-/bu-</i>	<i>anya</i>	<i>mù</i>	<i>àmu</i>	–

Note: \* forms vary tonally according to grammatical context.

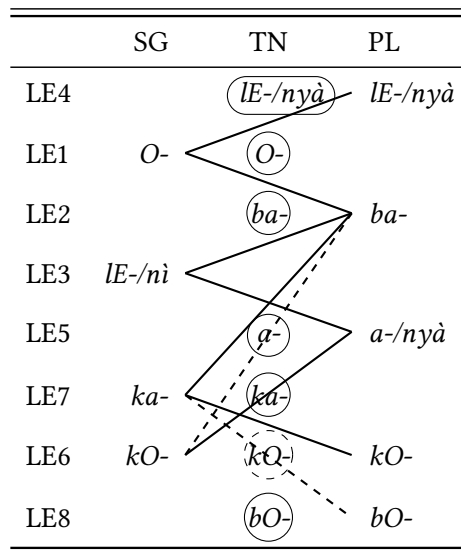


Figure 16: Gender system of Lelemi (based on Allan 1973)

Heine (1968: 114–115, 1982: 197–198) has also presented an analysis of the noun classification system of Lelemi with a focus on the Tetemang variety, which in turn has been reanalyzed by Corbett (1991: 173–175) from his typological perspective on gender. Figure 17 summarizes the results, including Corbett’s argument that some agreement class pairs should be viewed as inquorate genders.

	SG	TN	PL
LE1	<i>o-</i>		
LE2		?	<i>ba-</i>
LE3	<i>le-</i>		
LE4/5			<i>a-/(le-)</i>
LE7	<i>ka-</i>		
LE6	<i>ko-</i>		<i>ko-</i>
LE8	<i>bo-</i>		

Figure 17: Gender system of Lelemi (based on Heine 1968 and Corbett 1991)

The considerable divergence between the gender systems in the Figures 16 and 17 may be partly accounted for by dialect differences, given that Allan and Heine focused on Baglo and Tetemang, respectively. It is clear, however, that some differences are due to diverse analytical approaches. One crucial point is the identification of the additional plural LE4 for which Heine (1968: 115) also appears to present evidence with the demonstrative *-mε* but which Corbett (1991: 173) discards as a case of an overdifferentiated target. Another major difference in Heine’s analysis of Lelemi (albeit not in his family reconstruction, see §3.4.3) is the non-recognition of single-class genders, although there are some likely candidates, notably with LE8.

A final but important point regarding the previous analyses of Lelemi relates to the typologically oriented interpretation of the philological framework to Niger-Congo noun classification. That is, the description of Lelemi, couched by Heine (1968, 1982) in this tradition, misled Corbett (1991: 173–175) to a confusing analysis in that he calls the language’s genders inappropriately “agreement classes”. That the presentation of Niger-Congo data in particular causes such problem appears to be significant, because in general this author has applied his cross-linguistic approach successfully to a wide range of structurally diverse and complex gender systems.

<sup>9</sup>The tone marking in the table follows Allan’s (1973) transcription:  $\acute{V}$  high tone,  $V$  mid tone,  $\check{V}$  low tone.

Table 11: Nominal form class system of Lelemi (based on Allan 1973: 97–124)<sup>9</sup>

NF	Form(s)	Example(s)
Ø	–	SG <i>wɛwɛ</i> ‘dog’ TN <i>sika</i> ‘money’; <i>twifɔ̃</i> ‘Twi speaking person/people’
O-	ɔ-/u-	SG <i>ù-culi</i> ‘person’; <i>ɔ̃-gbà</i> ‘foot’ TN <i>ù-bòja</i> ‘blood’
BA-	ba-/be-	PL <i>bà-wɛwɛ</i> ‘dogs’; <i>bè-culi</i> ‘people’; <i>bè-kùkù</i> ‘owls’; <i>bè-se</i> ‘goats’; <i>bà-làkpi</i> ‘snakes’; <i>be-yu</i> ‘monkeys’
LE-	lɛ-/li-	SG <i>lì-kùkù</i> ‘owl’; <i>lɛ-nimì</i> ‘eye’ TN <i>lɛ-na</i> ‘meat’ PL <i>lè-gbà</i> ‘feet’
A-	a-/e-	SG <i>è-se</i> ‘goat’ TN <i>a-ba</i> ‘mud’ PL <i>a-nimì</i> ‘eyes’; <i>e-ji</i> ‘trees’
KO-	kɔ-/ku-	SG <i>kɔ̃-làkpi</i> ‘snake’; <i>ku-ji</i> ‘tree’ TN <i>ku-tu</i> ‘soup’ PL <i>kɔ̃-bwa</i> ‘hats’
KA-	ka-/ke-	SG <i>ke-yu</i> ‘monkey’; <i>kà-bwa</i> ‘hat’; <i>ke-mo</i> ‘farm’ TN <i>ka-na</i> ‘porridge’
N-	m-/n-/ɲ-	TN <i>n-tu</i> ‘water’; <i>ɲ-kpa</i> ‘life’ PL <i>m-mo</i> ‘farms’; <i>n-culi</i> ‘people (with NUM)’
BO-	bɔ-/bu-	TN <i>bɔ̃-ɲwa</i> ‘cooking’

Turning to Lelemi’s system of noun form and deriflection classes, Allan’s information can be summarized as in Table 11 and Figure 18.

Although Lelemi’s crossed gender system is already complex, its deriflection system is yet more elaborate, due notably to an additional prefixless nominal form class and another one in *N-*. It comprises 11 singular-plural affix pairings, albeit three of them inquorate. Nominal form classes are remarkable regarding their number behavior in that most of them are attested with more than one number value (only *BA-* and *BO-* are restricted to plural animates and transnumeral infinitives, respectively), and three of them are even attested in both singular and plural. Most of the discrepancies between gender and deriflection are thus due to the fact that agreement and nominal form classes show numerous patterns diverging from the expected biunique Niger-Congo canon, as shown in Figure 19.

## 5 Niger-Congo “noun classes” conflate gender with deriflection

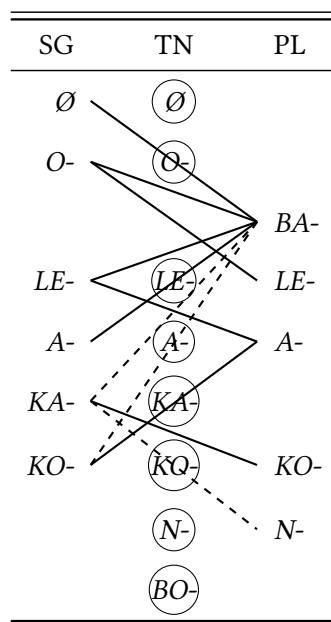


Figure 18: Deriflection system of Lelemi (based on Allan 1973: 100)

	AGR		NF	Number*
LE2	ba-	—	BA-	PL
	X	—	Ø	TN, SG, PL
LE1	O-	—	O-	TN, SG
LE5	a-	—	A-	TN, SG, PL
LE3	le-	—	LE-	SG
LE4	le-	—	X	TN, PL
LE6	ko-	—	KO-	TN, SG, PL
LE7	ka-	—	KA-	TN, SG
LE8	bo-	—	BO-	TN, (PL)
	X	—	N-	TN, PL

Note: X = no independent class counterpart in the other class type.

\* may join behavior for both AGR and NF

Figure 19: Mapping of agreement and nominal form classes in Lelemi (based on Allan 1973: 128)

### 3.4.3 Proto-Ghana-Togo-Mountain

The noun classification systems of Ghana-Togo-Mountain languages have been subject to historical-comparative analysis by Heine (1968). Since the very genealogical unity of the group is disputed, Heine's results are in principle controversial. In this context, however, we focus on another problem of his reconstruction, namely that he closely follows the problematic philological approach to Niger-Congo "noun classes", which obscures a transparent treatment of gender and nominal deriflection. Heine (1968: 112) writes:

Ein Nominalklassensystem liegt vor, wenn

- a) Nominalklassen bestehen, d.h. die Nomina durch Affixe in Klassen eingeteilt werden,
- b) Paarigkeit der Klassenaffixe vorhanden ist, d.h. einem sg-Affix ein bestimmtes pl-Affix entspricht bzw. umgekehrt, und wenn
- c) nach einer Nominalklassenkonkordanz verfahren wird, d.h. wenn den Nominalklassenaffixen an verschiedenen grammatischen Kategorien regelmäßig zugeordnete Klassen-Zeichen entsprechen.

[We speak of a noun class system if a) there are noun classes, that is, nouns are sorted by affixes into different classes; b) the class affixes occur in pairs, that is, a certain singular affix corresponds to a certain plural affix and vice versa; and if c) there is noun class concord, that is, if the noun class affixes correlate regularly with class exponents on different grammatical categories.]

Heine's awareness of the importance of agreement is reflected in his data presentation for single languages (ibid.: 113–123) as well as the exclusion of three languages from the reconstruction that according to him (ibid.: 276–277) no longer display class concord, namely Ikposo, Igo, and Animere (it turns out that this holds in fact only for the first language). Nevertheless, he focuses predominantly on the nominal affix system and often conflates agreement and noun forms, which makes it hard to distinguish the two. Finally, when reconstructing the "noun class" system of the entire group (ibid.: 187–211), he almost exclusively discusses the noun affixes; only in rare, unclear cases does he resort to the role of agreement forms.

A final point, which has also been made in §3.3 regarding the comparative work on Guang, concerns the reconstruction bias toward Proto-Bantu. Heine's proto-system, schematized in Figure 20, demonstrates that the inventory and numbering of the majority of his "noun classes" are, to the extent possible, clearly

	SG	TN	PL	
1/3	*o-		*ba- 2 *i- 4	
7	*ki-		*bi- 8	
5	*li-		*a- 6/10 *ku- 15	
9	*ku-			
13	*ka-		*bu- 14	
11		*N-		
12		*ti-		

Figure 20: “Noun class” system of Proto-Ghana-Togo-Mountain by Heine (1968: 187)

modeled on and also implicitly justified (ibid.: 187) by the conflated Proto-Bantu system, whose two components were shown in Figure 8 of §2.

Since Heine’s (1968) work many studies dealing to different degrees with the noun classification systems of individual Ghana-Togo-Mountain languages have appeared. Despite the much more complete data available today it remains hard to reconstruct a robust proto-system, irrespective of the classificatory status of the group. This is because most language-specific treatments are still biased toward nominal form classes and deriflections and neglect agreement, which is crucial for determining the gender system. That is, we have come across studies for only three of the 16 languages where the agreement and resulting gender systems receive primary attention by the respective authors, namely Zaske (2007) on Anii, Essegbey (2009) on Nyangbo, and Agbetsoamedo (2014a, 2014b) on Selee, while in all other descriptions this domain plays a secondary role, is overly conflated with nominal form classes, or is lacking altogether.

## 4 Summary

We have outlined the traditional approach to the noun categorization systems of the Niger-Congo type found in a large number of African languages and argued that it is in need of revision for the sake of better language-specific synchronic as

well as historical-comparative analyses. This holds in addition to the comparative bias toward the Bantu system, which tends to conceal a large part of the existing diversity across Niger-Congo languages.

One bias in the “noun class” framework is the strong focus on the affix status of class exponents. One consequence in the realm of nominal form classes is the overall analytical neglect of nouns without class affixes despite their important and partly diagnostic role in the nominal system.

Another crucial problem of the current Niger-Congo approach is the stereotypical view about agreement and nominal form classes in that the large majority of “noun classes” are assumed to be functionally dedicated to a specific gender and number value. As shown in the discussion of Proto-Bantu in §2, this situation is not even universal in the group that was the inspiration for this assumption. However, the degree of deviation from this hypothetical prototype can be much higher, so that this overgeneralized view should give way to a more neutral approach. In particular, this phenomenon throws a different light on the underlying number system in that the overall importance of transnumeral nouns seems to be higher than commonly assumed. That is, the data should no longer be dealt with according to a simple and universal singular-plural distinction.

The last and most important drawback of the traditional Niger-Congo framework is that its central concept of “noun class” conflates two independent linguistic phenomena associated with nouns: gender agreement between a nominal trigger and its target and deriflection reflected in morphological and/or phonological regularities of nouns. Their unified treatment has several negative effects for the current investigation of this domain. These are in particular an inappropriate focus on deriflection systems, a resulting neglect of a transparent and comprehensive analysis of agreement-based gender, and finally an impeded investigation of the exact relationship between the two distinct components, including their complex interdependency.

The disadvantages of the “noun class” concept negatively impact the transparency and even adequacy of language-specific descriptions. In the worst case, it may be impossible to establish the inventory of a language’s gender distinctions and its semantic and formal basis in spite of a lengthy treatment of “noun classes”. As discussed above, this is not restricted to a case like the heavily restructured Akan treated in §3.2, for which scholars go into great detail about its classificatory morphology on nouns but fail to explicitly identify the occasional existence of an animacy-based gender system.

Synchronic descriptive problems inevitably carry over to the historical reconstruction of noun classification in Niger-Congo, as shown for the Guang and Ghana-Togo-Mountain groups in §3.3 and §3.4, respectively. The general bias



toward the Bantu family aside, available proto-systems are not only unrealistic vis-à-vis the attested modern data but simply difficult to interpret linguistically in mixing distinct grammatical phenomena in a single paradigm.

Last but not least, it is hard to impossible for typologists to integrate a considerable amount of Niger-Congo data, in particular on complex systems, in cross-linguistic surveys on gender due to the intractable amalgamation of gender and deriflection. The typological incompatibility and thus “opaqueness” of many Niger-Congo descriptions deprives this research domain of interesting cases the analysis of which is necessary in order to arrive at meaningful cross-linguistic generalizations.

We venture that the cross-linguistic framework outlined in §1 is universally viable for language-specific, historical-comparative, and typological analyses. The restricted data presented here suggest several generalizations that are worth testing against a wider range of data. For example, the observation made in Güldemann (2000) that agreement classes need not be dedicated to specific gender and number values is demonstrably relevant for a much larger number of languages, and it can also be extended in Niger-Congo to nominal form classes. As proposed in Güldemann (2000), the degree of this functional insensitivity of classes is reflected in the ratio between genders and agreement classes (or, for that matter, between deriflections and nominal form classes). In typological comparison, this promises to serve as a good proxy for assessing basic structural differences between systems.

There is another conclusion that may turn out to be cross-linguistically significant, even though the data presented here are admittedly limited. That is, in languages with gender-sensitive noun morphology these deriflection systems are regularly more complex, or at least not simpler, than the associated gender systems in terms of inventory as well as systemic structure as per Heine (1982) and Corbett (1991).

For Niger-Congo languages, one can assume that the two subsystems of this nominal domain were originally very similar. This suggests for this group that deriflection systems tend to be more conservative than gender systems. With respect to the former, the transfer of individual or entire groups of nouns from one to another nominal form class, the merger of nominal form classes, and the resulting effects on deriflections are certainly rampant in the family. However, the changes in agreement-based gender marking are recurrently even more frequent and drastic, up to the reorganization, or even loss, of the entire system.

As long as the divergences between the two subsystems of gender and deriflection are minor, they will not differ dramatically in terms of their classification of nouns into sets. However, quite a few cases in Niger-Congo are differ-

ent. For example, Akan, dealt with in §3.2, possesses a binary system of animate vs. inanimate gender but an elaborate deriflection system with more and different categorizing distinctions. Languages of this type inform the new topic of so-called “concurrent systems” of noun classification, as investigated recently by Fedden & Corbett (2017) but for which the authors failed to recognize the relevance of Niger-Congo. Thus, a more detailed and typologically sound investigation of some of its languages where deriflection and gender have grown apart is a very worthwhile undertaking for the future.

In summary, this paper attempts to make two major contributions to the treatment of gender. First, the linguistic analysis of Niger-Congo-type noun classification systems should be better aligned with a sound cross-linguistic perspective. The detrimental philological approach, which is of a substantial rather than merely terminological nature, is not necessitated by any linguistic structures in Niger-Congo, however quirky they may appear from a cross-linguistic view. Second, we make a new proposal for a universally applicable framework for gender systems, especially useful if gender interacts intimately with the morpho(phono)logy of nouns. The approach based on the four analytical concepts outlined in §1 could not be fully expounded here by means of a wider language sample. However, its viability has been shown for the specific gender-system profile of the important group of Niger-Congo languages. It has also been applied successfully to structurally quite different languages from such families as Kx’a and Tuu in southern Africa, Kadu and Cushitic in northeastern Africa, and yet others. Hence, we venture to review the approach to gender from a wider typological perspective in line with the present framework.

## **Acknowledgments**

This paper or parts thereof were previously presented at the International Workshop on “Grammatical Gender and Linguistic Complexity” at the Department of Linguistics of Stockholm University, 20–21 November 2015; at the Linguistics Department of the University of California Berkeley, 30 March 2016; the International Conference “Toward Proto-Niger-Congo II” at LLACAN, Paris, 1–3 September 2016; at the Department of African and Ethiopian Studies of Hamburg University, 24 January 2018; and at the Diversity Linguistics Seminar at Leipzig University, 1 February 2018. We are grateful for the fruitful feedback by the respective audiences. Our thanks also go to the extensive and productive comments by the editors of this volume, two anonymous reviewers, and Martin Haspelmath. Last but not least, we gratefully acknowledge the funding received

from the German Research Foundation (DFG) for the project ‘Noun classification in Africa between gender and declension’ within which the greater part of our research presented here was carried out.

## Special abbreviations

The following abbreviations are not found in the Leipzig Glossing Rules:

AGR	agreement class	NUM	numeral
AN	animate	PART	participle
CONC	pronominal concord	PERF	perfect
D	distal	PRO	pronoun
IAN	inanimate	TN	transnumeral
NF	nominal form class		

Arabic numbers represent agreement classes while Roman numbers represent genders.

## References

- Abiodun, Michael A. 1989. The class system of Iggede nouns. *Journal of West African Languages* 19(2). 51–56.
- Agbetsoamedo, Yvonne. 2014a. *Aspects of the grammar and lexicon of Selee*. Stockholm University dissertation.
- Agbetsoamedo, Yvonne. 2014b. Noun classes in Selee. *Journal of West African Languages* 41(1). 95–124.
- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Allan, Edward Jay. 1973. *A grammar of Buem: The Lelemi language*. School of Oriental & African Studies, University of London dissertation.
- Babou, Cheikh Anta & Michele Loporcaro. 2016. Noun classes and grammatical gender in Wolof. *Journal of African Languages and Linguistics* 37(1). 1–57.
- Blench, Roger. 2009. Do the Ghana-Togo Mountain languages constitute a genetic group? *Journal of West African Languages* 36(1/2). 19–35.
- Bodomo, Adams & Charles Marfo. 2006. The morphophonology of noun classes in Dagaare and Akan. *Studi Linguistici e Filologici Online* 4(2). 205–243.
- Bokula, François-Xavier. 1971. Formes nominales et pronominales en mba. *Africana linguistica* 5. 41–77.

- Borchardt, Nadine. 2011. *The numeral system of Ikaan, a Benue-Congo language of Nigeria* (Asien- und Afrika-Studien der Humboldt-Universität zu Berlin 37). Wiesbaden: Harrassowitz.
- Canu, Gaston. 1976. *La langue mò:re, dialecte de Ouagadougou (Haute-Volta): Description synchronique* (Langues et Civilisations à Tradition Orale 16). Paris: SELAF.
- Christaller, Johann Gottlieb. 1875. *A grammar of the Asante and Fante language called Tshi (Chwee, Twi): Based on the Akuapem dialect with reference to the other (Akan and Fante) dialects*. Basel: Basel Evangelical Missionary Society.
- Connell, Bruce A. 1987. Noun classification in Lower Cross. *Journal of West African Languages* 17(1). 110–125.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2013. Number of genders. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/30>.
- de Wolf, Paul P. 1971. *The noun class system of Proto-Benue-Congo* (Janua Linguarum, Series Practica 167). The Hague/Paris: Mouton.
- Di Garbo, Francesca. 2014. *Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of Africa*. Stockholm University dissertation.
- Dolphyne, Florence Abena & Mary Esther Kropp Dakubu. 1988. The Volta-Comoe languages. In Mary Esther Kropp Dakubu (ed.), *The languages of Ghana*, 50–90. London: Kegan Paul International (for the International African Institute).
- Doneux, Jean L. 1975. Hypothèses pour la comparative des langues atlantiques. In *Africana Linguistica* 6 (Annalen Wetenschappen van de Mens 88), 41–129. Tervuren: Koninklijk Museum voor Midden-Afrika.
- Elugbe, Ben O. 1983. Noun class prefixes in Proto-Edoid. In Ivan R. Dihoff (ed.), *Current approaches to African linguistics* (Publications in African Languages and Linguistics 1), 59–83. Dordrecht/Cinnaminson: Foris.
- Essegbey, James. 2009. Noun classes in Tutrugbu. *Journal of West African Languages* 36(1/2). 37–56.
- Evans, Nicholas. 1997. Head classes and agreement classes in the Mayali dialect chain. In Mark Harvey & Nicholas Reid (eds.), *Nominal classification in Aboriginal Australia*, 105–146. Amsterdam: John Benjamins.
- Evans, Nicholas, Dunstan Brown & Greville G. Corbett. 1998. Emu divorce: A unified account of gender and noun class assignment in Mayali. In M. Catherine Gruber, Derrick Higgins, Kenneth S. Olson & Tamra Wysocki (eds.), *Pa-*

- pers from the 34th regional meeting of the Chicago Linguistic Society, 147–172. Chicago: CLS.
- Fedden, Sebastian & Greville G. Corbett. 2017. Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa: A journal of general linguistics* 2(1). 1–47. DOI:10.5334/gjgl.177
- Good, Jeff. 2012. How to become a “Kwa” noun. *Morphology* 22(2). 293–335.
- Güldemann, Tom. 2000. Noun categorization systems in non-Khoe lineages of Khoisan. *Afrikanistische Arbeitspapiere* 63. 5–33.
- Güldemann, Tom. 2018. Historical linguistics and genealogical language classification in Africa. In Tom Güldemann (ed.), *The languages and linguistics of Africa* (The World of Linguistics 11), 58–444. Berlin: De Gruyter Mouton.
- Güldemann, Tom & Ines Fiedler. 2016. What gender systems? Agreement classes vs. noun form classes in Niger-Congo with particular reference to Ghana-Togo-Mountain languages. Paper presented at the Linguistics Department, University of California Berkeley, 30 March 2016.
- Güldemann, Tom & Samuele Maniscalco. 2015. A new approach to gender in Somali. Paper presented at the international workshop “Grammatical gender and linguistic complexity”, Stockholm University, 20–21 November 2015. [http://www.ling.su.se/polopoly%5C\\_fs/1.266854.1453731742!/menu/standard/file/Gueldemann%5C\\_Somali%5C%20Gender%5C%20H.pdf](http://www.ling.su.se/polopoly%5C_fs/1.266854.1453731742!/menu/standard/file/Gueldemann%5C_Somali%5C%20Gender%5C%20H.pdf).
- Guthrie, Malcolm. 1948. Gender, number and person in Bantu languages. *Bulletin of the School of Oriental and African Studies* 12(3/4). 847–856.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath (eds.). 2018. *Glottolog 3.2*. Jena: Max Planck Institute for the Science of Human History. <http://glottolog.org/>.
- Hansford, Keir Lewis. 1990. *A grammar of Chumburung: A structure-function hierarchical description of the syntax of a Ghanaian language*. University of London dissertation.
- Haspelmath, Martin. 1989. From purposive to infinitive – a universal path of grammaticization. *Folia Linguistica Historica* 10(1/2). 287–310.
- Heine, Bernd. 1968. *Die Verbreitung und Gliederung der Togorestsprachen* (Kölner Beiträge zur Afrikanistik 1). Berlin: Reimer.
- Heine, Bernd. 1982. African noun class systems. In Hansjakob Seiler & Christian Lehmann (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen, Teil I: Bereich und Ordnung der Phänomene*, 189–216. Tübingen: Narr.
- Hyman, Larry M. 1980. *Noun classes in the Grassfields Bantu borderland* (Southern California Occasional Papers in Linguistics 8). Los Angeles: Department of Linguistics, University of Southern California.

- Kilarski, Marcin. 2013. *Nominal classification: A history of its study from the classical period to the present*. Amsterdam: John Benjamins.
- Manessy, Gabriel. 1967. Évolution de la classification nominale dans les langues gurunsi (group voltaïque). In Gabriel Manessy (ed.), *La classification nominale dans les langues négro-africaines*, 207–224. Paris: CNRS.
- Manessy, Gabriel. 1975. *Les langues oti-volta: Classification généalogique d'un groupe de langues voltaïques* (Langues et Civilisations à Tradition Orale 15). Paris: SELAF.
- Manessy, Gabriel. 1987. La classification nominale en proto-guang. *Afrikanistische Arbeitspapiere* 9. 5–49.
- Marchese, Lynell. 1988. Noun classes and agreement systems in Kru: A historical approach. In Michael Barlow & Charles A. Ferguson (eds.), *Agreement in natural languages: Approaches, theory, descriptions*, 323–341. Stanford: CSLI.
- Meeussen, Achille E. 1967. Bantu grammatical reconstructions. *Africana Linguistica* 3(1). 79–121.
- Miehe, Gudrun. forthcoming. Noun class systems in Gur and Plateau languages: Comparing the diachronic developments. In Theda Schumann & Roland Kießling (eds.), *Endangered languages in contact: The Plateau languages of Nigeria*. Cologne: Köppe.
- Miehe, Gudrun, Ulrich Kleinewillinghöfer, Manfred von Roncador & Kerstin Winkelmann. 2012. Overview of noun classes in Gur (II) (revised and enlarged version). In Gudrun Miehe, Brigitte Reineke & Kerstin Winkelmann (eds.), *Noun class systems in Gur languages. Vol. 2: Overview of noun classes in Gur (II) (revised and enlarged version): North Central Gur languages*, 5–37. Cologne: Köppe.
- Neuhaus, Simon. 2008. Das Genussystem des Krongo. Lizenzat thesis, Universität Zürich.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Osam, Emmanuel Kweku. 1993. The loss of the noun class system in Akan. *Acta Linguistica Hafniensia* 26. 81–106.
- Painter, Colin. 1970. *Gonja: A phonological and grammatical study*. Bloomington: Indiana University Press.
- Pasch, Helma. 1986. *Die Mba-Sprachen: Die Nominalklassensysteme und die genetische Gliederung einer Gruppe von Ubangi-Sprachen* (Sprache und Geschichte in Afrika 6). Hamburg: Buske.
- Reineke, Bigitte. 1966. Die nominale Struktur des Nkunya. *Mitteilungen des Instituts für Orientforschung* 3(12). 209–219.

- Schadeberg, Thilo. 1981a. *A survey of Kordofanian*. Vol. 1: *The Heiban group*. Hamburg: Buske.
- Schadeberg, Thilo. 1981b. *A survey of Kordofanian*. Vol. 2: *The Talodi group*. Hamburg: Buske.
- Snider, Keith. 1988. The noun class system of Proto-Guang and its implications for internal classification. *Journal of African Languages and Linguistics* 10(2). 137–164.
- Sterk, Jan P. 1976. The Hohumono noun class system. *Journal of West African languages* 11(1/2). 119–140.
- Sterk, Jan P. 1978. The noun class system of Gade (Nigeria). *African Languages* 4. 24–43.
- Stewart, John M. 2002. The potential of Proto-Potou-Akanic-Bantu as a pilot Proto-Niger-Congo, and the reconstructions updated. *Journal of African Languages and Linguistics* 23(2). 197–224.
- Van de Velde, Mark L. O. 2006. Multifunctional agreement patterns in Bantu and the possibility of genderless nouns. *Linguistic Typology* 10(2). 183–221.
- Voll, Rebecca M. 2017. *A grammar of Mundabli: A Bantoid (Yemne-Kimbi) language of Cameroon*. Leiden University dissertation.
- Voorhoeve, Jan & Paul P. de Wolf (eds.). 1969. *Benue-Congo noun class systems*. Leiden: Afrika-Studiecentrum for the West African Linguistic Society.
- Welmers, William E. 1971. The typology of the Proto-Niger-Kordofanian noun class system. In Kim Chin-Wu & Herbert Stahlke (eds.), *Papers in African linguistics*, 1–16. Edmonton/Champaign: Linguistic research.
- Welmers, William E. 1973. *African language structures*. Berkeley: University of California Press.
- Westermann, Diedrich. 1922. *Die Sprache der Guang in Togo und auf der Goldküste und fünf andere Togosprachen*. Berlin: Reimer.
- Westermann, Diedrich. 1935. Nominalklassen in westafrikanischen Klassensprachen und in Bantusprachen. *Mitteilungen des Seminars für Orientalische Sprachen zu Berlin* 38. 1–53.
- Zaliznjak, Andrej A. 1964. K voprosu o grammatičeskix kategorijax roda i oduševlennosti v sovremennom russkom jazyke. *Voprosy jazykoznanija* 4. 25–40.
- Zaske, Stephanie G. 2007. *N’ushile ma – apprentissage de la langue anii*. Bassila: Summer Institute of Linguistics (SIL), Togo/Bénin.





## Chapter 6

# Gender in Uduk

Don Killian

University of Helsinki

Uduk, a Koman language spoken on the border of Ethiopia and Sudan, evinces a number of unusual characteristics in its system of gender marking. Uduk has two gender classes, with agreement displayed primarily in the verbal system and adjacent case-marking particles. In contrast to related Koman languages, however, semantics play a minimal role in class assignment, unrelated to biological sex. Furthermore, as biological sex does not play a role in gender assignment in general, personal pronouns do not differentiate gender in any person. Instead, all personal pronouns are assigned to Class 1 in the same manner that nouns would be. Lastly, Uduk shows some unorthodox aspects in the way it indexes gender on verbs, using what might be considered subtractive morphology.

This article looks at the complexity and features of gender in Uduk from a typological perspective; despite some unorthodox and atypical typological features, however, the system does not appear to be complex.

**Keywords:** Uduk, gender, assignment, Koman, adjacency, ditropic.

## 1 Background

Koman languages form a small language family spoken along the borderland area of Ethiopia, Sudan and South Sudan. The family is comprised of four living languages: Gwama (Kwama) [kmq], T'apo (also known as Opo or Opuo) [lgn], Komo [xom] and Uduk (Tw'ampa) [udu]. A fifth language which is now extinct, Gule, was placed into Koman by Greenberg with relatively little data available (Greenberg 1963), and its placement in Koman is tentative.

The presence of gender distinctions on pronouns in Koman languages was noted early on, but no research until recently has uncovered any signs of a nominal grammatical gender system, which all extant Koman languages have in some



fashion.<sup>1</sup> The data on Uduk presented here is based on thirteen months of field-work between 2011 and 2014 in Ethiopia.

## 2 Introduction

Gender is a noun classification strategy in which nouns are encoded to belong to a particular lexical class, which is further “reflected in the behavior of associated words” (Hockett 1958: 231). This is commonly referred to as *agreement*, a relationship in which one element takes an inflectional form determined by semantic or morphosyntactic properties of another element. Following Corbett (2006), the element which determines the agreement is the *controller*, and the element whose form is determined by agreement is the *target*.

As the notion of agreement implies that the controller is present (cf. Corbett 2006), the term *indexation* is used instead of agreement. Indexation is defined here as the morphosyntactic realization of a controller’s capacity to control a target, with the controller being either present or recoverable or identifiable in some way. This may be done inflectionally through means of an affix or clitic, but this may also occur on a broader level by use of particular constructions, as Uduk does not always index gender on targets through inflectional markers. In particular, when in object position, one class of nouns actually constrains verb paradigms, limiting the possible *subject* cross-referencing markers on the verb. Thus, it is possible to determine the gender of the object from the morphology of the verb, despite there being no affix on the verb expressing gender agreement with the object.

Many other aspects of the Uduk gender system show themselves to be unorthodox in nature. Semantic assignment exists only for a very small part of the lexicon, formal assignment (in terms of word formation rules) for another very small part, with the rest being largely arbitrary. Semantics in general play a smaller role than usual in gender assignment, and Uduk’s cut-off point in the animacy hierarchy for semantic assignment is higher than simply ‘human’.

Furthermore, typical indexation targets of gender cross-linguistically include demonstratives, determiners, personal pronouns, relative pronouns, adjectives and verbs (Di Garbo 2014). For Uduk, the only target in this list is verbs. In addition to verbs, indexation is primarily indicated on a single clitic or particle which immediately precedes the controller, and on prepositions.

---

<sup>1</sup>The Yabus dialect of Uduk appears to be an exception to this, and does not have any grammatical gender.

It is worth considering Uduk's gender system in terms of its linguistic complexity.<sup>2</sup>

For some principles governing local complexity, see Di Garbo (2016: 50) or Audring (2019 [this volume], §2.3). In addition to those metrics, there are at least two factors which may play a role, arbitrariness and adjacency, although how they fit precisely remains to be determined. Complexity is discussed further in §5.

### 3 Introduction to gender in Uduk

All nouns in Uduk, including proper nouns, are allocated to one of two possible grammatical gender classes, labeled *Class 1* and *Class 2*.

Gender in Uduk is covert, and not marked directly on nouns. Gender distinctions are seen most commonly through the presence or absence of the Class 2 clitic *à=*;<sup>3</sup> this marker, however, is optional when the noun occurs in isolation. Furthermore, if gender is indexed on a previous word in the phrase, then *à=* is not used with the noun. Vocative use also neutralizes gender distinctions in many instances. When directly addressing an individual, all personal names<sup>4</sup> and most Class 2 kinship terms remove *à=*; a handful of kinship terms may retain *à=* to indicate a type of informality. In all other known instances, Class 2 nouns occur preceded by *à=*.

Gender indexation primarily occurs on case-marking clitics or particles which immediately precede the controller. Prepositions, conjunctions, and complementizers also undergo a simple phonological alternation, depending on the gender of the noun that follows, and verbs also vary in their conjugation paradigms depending on the gender of a postverbal object. In some instances, clitics may be considered ditropic clitics,<sup>5</sup> phonologically attaching to the constituent which immediately precedes the clitic. However, unlike more typical situations of ditropic clitics, phonological hosts are more constrained in Uduk. Further details are discussed in §3.2 below after a general introduction to grammatical relations in Uduk.

<sup>2</sup>Linguistic complexity refers here to the amount of information needed to describe the system, following e.g. Dahl (2004) and Miestamo (2008).

<sup>3</sup>Transcriptions used here follow the IPA, except for <y>, which represents IPA *j*, and <j>, which represents IPA *ɟ*.

<sup>4</sup>All personal names are assigned to Class 2, discussed in more detail in §4.

<sup>5</sup>Ditropic clitics are a type of clitic which occur before a particular lexical class or syntactic phrase functionally related to the clitic in question, but the clitics nonetheless phonologically attach to the constituent on the 'other' side instead. This host generally is structurally and functionally highly variable, and shows little functional relation to the clitic. For more details, see Cysouw (2005).

### 3.1 Grammatical relations overview

Case and constituent order are intertwined in Uduk, and it is not possible to discuss one without the other. The order of constituents frequently changes, and the order of the arguments affects the way in which these are encoded.<sup>6</sup>

Uduk follows a verb-second pattern similar to that of some neighboring Nilotic languages. Intransitive clauses primarily use SV order, with occasional instances of VS order in specific types of subordinate clauses. Transitive clauses regularly alternate between OVA and AVO, and cannot be easily characterized as having a dominant constituent order. Other constituent orders do not occur in main clauses.

The only situation in which an argument triggers the presence of morphological case marking is when it occurs in the position immediately following the verb. Other core relations are not case-marked, irrespective of whether they occur before or after the immediately postverbal position. If the postverbal argument is O, this may be indicated by an Accusative ditropic clitic which phonologically attaches onto the verb. If the argument is A, the verb is marked by a ditropic clitic indicating Ergative case.<sup>7</sup> Note that verbs ending in vowels add a nasal suffix if the argument that follows is marked with Ergative case.

Table 1 shows the different case markers used in Uduk.<sup>8</sup> All case-marking enclitics are ditropic.

Some examples are as follows:<sup>9</sup>

---

<sup>6</sup>The framework used here to refer to argument structure is based on a division elaborated on by Dixon (1994), in which participants of a clause are divided into core and peripheral roles. Core functions include the transitive subject (A), the intransitive subject (S), and the transitive object (O); all other participants are treated as peripheral.

<sup>7</sup>The Ergative case primarily indicates the subject of a transitive clause; however, in two instances, namely relative clauses and temporal adverbial subordinate clauses, the same marker is also used with subjects of intransitive clauses as well. In these two clause types, then, Uduk would be considered as having Marked Nominative case marking rather than Ergative. All Marked Nominative examples are nonetheless glossed as *ERG*, however, to simplify matters. For further details, see Killian (2015).

<sup>8</sup>Absolutive is not used here to refer to a case encompassing S and O, but is used in a more general sense to refer to most situations in which the noun is not marked for Accusative, Associative, Ergative, or Genitive. This includes all preverbal arguments and second arguments after the verb in ditransitive constructions. Absolutive Class 2 *ā=* is not used in prepositional phrases, however, and optionally in citation form. Associative is used to refer to a type of noun-noun collocation in which the second noun modifies the first in some way, typically conveying either possession or association. It is similar to the Genitive, but the relationship between the two nouns in the Associative is much broader and less defined. For further details, see Killian (2015).

<sup>9</sup>The underlined argument indicates the topical argument of a transitive clause.

Table 1: Case Markers

	(ABS)	ACC	ASSOC	ERG	GEN
Class 1	∅	∅	∅	= $\bar{a}$	$g\dot{i}$
Class 2	$\dot{a}=$	= $\bar{a}$	= $\bar{a}$	= $m\bar{a}$	= $m\bar{a}$

## (1) Intransitive

$\dot{a}=c'i$        $k'\dot{u}t^h-\dot{u}d'$   
 CL2=child(CL2) cough:IPFV-3SG  
 ‘The child coughed.’

## (2) Transitive, AVO order

$\dot{a}=p\acute{a}w$        $\bar{u}r-\dot{u}d'=\bar{a}$        $t'ik^h$   
 CL2=cat(CL2) chase:IPFV-3SG=ACC.CL2 rat(CL2)  
 ‘The cat chased the rat.’

## (3) Transitive, OVA order

$\dot{a}=p\acute{a}w$        $w\acute{u}c'=m\bar{a}$        $k'\acute{a}$   
 CL2=cat(CL2) bite:PFV=ERG.CL2 dog(CL2)  
 ‘The dog bit the cat.’

### 3.2 Gender and case marking

As mentioned in the previous section, gender differentiations are found in case marking. Uduk encodes gender and case marking cumulatively, with a single combined morph to represent multiple features. Case is generally marked by clitics or particles immediately preceding the noun, and case markers which indicate core arguments only occur in the immediately postverbal position.

All case markers except Class 2 Absolutive  $\dot{a}=$  and Class 1 Genitive  $g\dot{i}$  are ditropic clitics, clitics which form phonological units with the immediately preceding element. Not all markers, however, are as bound as others, and boundedness forms something of a continuum.

Accusative Class 2  $=\bar{a}$  and Ergative Class 1  $=\bar{a}$  both form relatively tight-knit phonological units with the verb, and trigger morphophonological changes on

the verb.<sup>10</sup> If a verb ends in a vowel, however, Accusative =*ā* does behave slightly differently compared to the Ergative =*ā*. Verbs ending in a vowel always add an extra -*n* to the end when occurring before Ergative case markers of either class, before Class 1 =*ā* as well as before Class 2 =*mā*. Accusative Class 2 =*ā* on the other hand simply attaches to whatever the final consonant or vowel is, including other vowels. Associative Class 2 =*ā* behaves identically to Accusative phonologically, but attaches to a noun rather than a verb.<sup>11</sup>

All case markers discussed except for Genitive Class 1 *gi* undergo phonological tonal alternations depending on the immediately preceding tone. This includes Accusative Class 2 =*ā*, Associative Class 2 =*ā*, Ergative Class 1 =*ā*, Ergative Class 2 =*mā*, and Genitive Class 2 =*mā*. The base tone of the case marker is mid, but lowers to low when immediately following a low tone. Neither Ergative Class 2 =*mā* nor Genitive Class 2 =*mā* trigger morphophonological changes, however.

Genitive Class 1 *gi* is not a clitic, but rather an independent particle which does not change tone or affect any consonants or tones around it.

Some simple examples of each form are given below.<sup>12</sup>

- (4) Accusative, Class 2

*k'wānī*      *lōb-ón=ā*      *k<sup>h</sup>úrā*  
 people(CL1) play:IPFV-3PL=ACC.CL2 ball(CL2)  
 'The people are playing football.'

- (5) Ergative, Class 1

*à=k<sup>h</sup>úrā*      *lōb=ā*      *k'wānī*  
 CL2=ball(CL2) play:IPFV=ERG.CL1 people(CL1)  
 'The people are playing football.'

- (6) Ergative Class 2

*à=k<sup>h</sup>úrā*      *lōb=mā*      *c'í*  
 CL2=ball(CL2) play:IPFV=ERG.CL2 child(CL2)  
 'The child is playing football.'

<sup>10</sup>Glottalized consonants in word-final position are unreleased. If any affixes or clitics are placed after them, they undergo a morphophonological alternation described in more detail in Killian (2015: 48).

<sup>11</sup>If the first noun in the Associative construction ends in a vowel and the consonant of the second noun begins with a plosive, a homorganic nasal is used in place of *ā*. For more details, see Killian (2015: 89).

<sup>12</sup>Clauses with Class 1 postverbal objects are not included, as they are a special case discussed in §3.5 below.

- (7) Genitive, Class 1  
*à=nós              gí              wàtí?*  
 CL2=pot(CL2) GEN.CL1 man(CL1)  
 ‘the man’s pot’
- (8) Genitive, Class 2  
*à=nós=mā                      bóm*  
 CL2=pot(CL2)=GEN.CL2 woman(CL2)  
 ‘the woman’s pot’
- (9) Associative, Class 1  
*à=rīs                      k’wānī*  
 CL2=many.PL(CL2) people(CL1)  
 ‘very many people’
- (10) Associative, Class 2  
*à=rīs=ā                      kúnù?*  
 CL2=many.PL(CL2)=ASSOC.CL2 owl(CL2)  
 ‘very many owls’

### 3.3 Prepositions, conjunctions, and complementizers

In addition to case marking, gender is also marked on prepositions, conjunctions, and complementizers in Uduk through a simple phonological alternation. If a preposition ends in *i*, this changes to *a* before Class 2 nouns, retaining the tone of the original vowel. If a preposition ends in a consonant or another vowel than *i*, then *a* attaches to the end of the preposition. As mentioned previously, if gender is marked on the previous element, then Class 2 marker *à=* is not used.

These alternations are likely based on a type of cliticization similar to case markers, but slightly more grammaticalized. Nonetheless, in occasional careful speech with *ḡàlì* ‘and, but’ for instance, it is possible to hear *ḡàlì à* before Class 2 nouns instead of *ḡàlā*.<sup>13</sup>

- (11) *ràk<sup>h</sup>              tā-∅              kúf      mò í              mīs*  
 cloud(CL1) COP:PFV-3SG white MO LOC:CL1 sky(CL1)  
 ‘The clouds are white in the sky.’

<sup>13</sup>Note that in the following examples, ‘zero clitics’ =∅ have been added to facilitate understanding.

- (12) *áhā wòl-á=ø yídé á k<sup>h</sup>òs*  
 1SG(CL1) pour:IPFV-1SG=CL1 water(CL1) LOC:CL2 cup(CL2)  
 ‘I poured the water in the cup.’
- (13) *é gǎm-ø=ø tō yán p’éní máná?*  
 2SG(CL1) find:IPFV-3SG=CL1 thing(CL1) DEM.PROX from:CL1 where(CL1)  
 ‘Where did you get this thing from?’
- (14) *gǎm-kā? p’éná Yúsif*  
 find:IPFV-ERG.1SG from:CL2 Yousef(CL2)  
 ‘I got (it) from Yousef.’

Predicative possession constructions also index the gender of the possessed noun on a preposition-like marker. These predicative possessive constructions are formed with the copula *tā* along with the particle *gi*, which becomes *gà* before Class II nouns (unlike Genitive *gi*, which becomes =*mā* before Class II nouns).

- (15) *wàtí? tā gi mì*  
 man(CL1) COP:PFV PP.CL1 goat(CL1)  
 ‘The man has a goat.’
- (16) *áhā tā-ná gà k’á*  
 1SG(CL1) COP:PFV-1SG PP.CL2 dog(CL2)  
 ‘I have a dog.’

Conjunctions and complementizers are preposition-like words used to connect clauses or phrases. Similar to prepositions, the gender of the immediately following word is marked on the conjunction or complementizer by an alternation of *i* to *a* for words ending in *i*, or by adding *a* to the end of words which end in consonants or vowels other than *i*.

The most frequent of these is *kí*, or *ká* for Class 2 nouns. It is a general complementizer which occurs with many different types of complement phrases and clauses, as well as subordinate clauses.

- (17) *áhā t<sup>h</sup>ōf-á kí wàtí? mĩ-d’=i*  
 1SG(CL1) think:IPFV-1SG COMP:CL1 man(CL1) do.AUX:IPFV:AD2-3SG=LNK  
*t’ā kí p<sup>h</sup>úd mò fwàné?*  
 CF.AUX COMP arrive MO today  
 ‘I thought that the man would have arrived today.’



- (18) *áhā t<sup>h</sup>ōf-á ká fōk' mī-d=i hét'*  
 1SG(CL1) think:IPFV-1SG COMP:CL2 rain(CL2) do.AUX:IPFV-3SG=LNK rain<sub>verb</sub>  
*kāt'ámō*  
 tomorrow  
 'I hope it rains tomorrow.'

With some adverbial phrase constructions, *kī* and *kā* with mid tones are used instead of *kí* and *ká* with high tones.

- (19) *únī dōf-ón kī mīs*  
 3PL(CL1) stand:IPFV-3PL with:CL1 sky(CL1)  
 'They stood up.'
- (20) (Beam & Cridland 1970)  
*jàmàs būnī k'ō-n kā ris*  
 kind(CL1) POSS.3PL exist.PL:PFV-3PL with:CL2 many(CL2)  
 'There are many kinds of them.'

There are three additional subordinating conjunctions: *wāk<sup>h</sup>kí* for conditional clauses, *gòm* for reason and adversative clauses, and *mèd* for temporal clauses. All of these alternate according to the gender of the noun which follows in the manner described above.

- (21) *wāk<sup>h</sup>kí wātí? k'óf-ód=ā shēt<sup>h</sup>, k'úp<sup>h</sup>*  
 if:CL1 man(CL1) kill:PFV-3SG=ACC.CL2 antelope(CL2), head(CL1)  
*tō mí-nù mí=i k<sup>h</sup>ál bwày cōm=á?*  
 thing(CL1) do.AUX:PFV-IMPRS do.AUX=LNK carry to:CL1 his.father(CL1)=Q  
 'If a person kills an antelope, is the head carried to the father's home?'
- (22) *wāk<sup>h</sup>ká c'í p<sup>h</sup>úd-úd mò yìl k'úmèd pé kwārā áw*  
 if:CL2 child(CL2) reach:IPFV-3SG MO year(CL1) thirteen or  
*k'úmèd í pé sú? ádī k'í t<sup>h</sup>él mí p'én=i máf mò*  
 twelve 3SG(CL1) NARR begin do.PART behind.PART=LNK marry MO  
 'If the child reaches the year thirteen or twelve then he can start to get married.'

The only native coordinating conjunction is *ḍàlì* (Class 2 *ḍàlà*) ‘and; but’, and it is very frequent.<sup>14</sup> It may coordinate clauses, noun phrases, and nouns.

- (23) *ḍàlì tōnt<sup>h</sup>é? yĩsā dì-d yĩsā=yà*  
**and:CL1** food(CL1) NEG exist.SG:PFV-3SG NEG=NEG  
 ‘And there was no food.’
- (24) (James 1979, The Birapinya Tree)  
*ḍàlà bóm nǎŋ-ø=ø gùb fēmēn*  
**and:CL2** woman(CL2) build:IPFV-3SG=ACC.CL1 house(CL1) alongside:CL1  
*bwày*  
 road(CL1)  
 ‘And a woman had built her house alongside the road.’

### 3.4 Prenominal modifiers

Out of all the prenominal modifiers, two of them index the gender of the noun they modify, namely the diminutive *ārí* and its irregular plural form *ũĩ*. Both the singular as well as the plural diminutive are lexically nouns themselves, with inherent gender (Class 1). However, they alternate their final vowel according to the gender of the following noun: *í* before Class 1, and *á* before Class 2.

- (25) *áhā mĩf-á=ø ārí mĩ*  
 1SG(CL1) see:IPFV-1SG=ACC.CL1 DIM:CL1(CL1) goat(CL1)  
 ‘I saw the little goat.’
- (26) *áhā mĩf-á=ø ārá nǎw*  
 1SG see:IPFV-1SG=ACC.CL1 DIM:CL2(CL1) cat(CL2)  
 ‘I saw the little cat.’

There is one special case in regards to prenominal modifiers that should also be mentioned, one of the only instances of non-adjacent indexation of gender. When prenominal modifiers modify a postverbal A argument, the verb does not agree with the inherent gender of the modifier, but rather with the noun that the prenominal modifier is modifying.

<sup>14</sup>Two other conjunctions borrowed from Arabic also exist: *wàlà* and *áw*, both meaning ‘or (used to rephrase something)’. Neither term alternates according to the gender of the noun which follows.

- (27) Class I Noun

à=bóm                      míf=à                      wàtí?  
 CL2=woman(CL2) see:IPFV=ERG.CL1 man(CL1)  
 ‘The man sees the woman.’

- (28) Class I Modifier, Class I Noun

à=bóm                      míf=à                      ḡàn                      wàtí?  
 CL2=woman(CL2) see:IPFV=ERG.CL1 big(CL1) man(CL1)  
 ‘The big man sees the woman.’

- (29) Class II Noun

wàtí?                      míf=mà                      bóm  
 man(CL1) see:IPFV=ERG.CL2 woman(CL2)  
 ‘The woman sees the man.’

- (30) Class I Modifier, Class II Noun

wàtí?                      míf=mà                      ḡàn=à                      bóm  
 man(CL1) see:IPFV=ERG.CL2 big(CL1)=ASSOC.CL2 woman(CL2)  
 ‘The big woman sees the man.’

Constructions of this type have only appeared in elicited circumstances, however, and speakers appeared to be somewhat reluctant to use them. Not all Uduk speakers would necessarily find these grammatical; many would find them odd, at the very least, and would avoid using postverbal A arguments with prenominal modifiers.

### 3.5 Verbs

Finite verbs are the last target for gender indexation presented here; verbs indicate the gender of O arguments in a rather unusual fashion.

In constructions in which the O argument is Class 2 (e.g. marked with the Accusative), the A argument is cross-referenced in the same way that S would be in monovalent clauses. Verbs with a 3SG subject are marked with  $-(V)d$ , and verbs with a 2SG, 2PL, or 3PL subject are marked with  $-(V)n$  on the verb. Verbs with 1SG and 1PL.EX subjects take  $-á$ , and 1PL.IN subjects take  $-à$ .

- (31) Class 2 O, 3SG subject

*wàtí? c'ít'-íd=ā yíd*  
 man(CL1) cut:IPFV-3SG=ACC.CL2 skin(CL2)  
 'The man is cutting the pelt.'

- (32) Class 2 O, 2SG subject

*é gām-án=ā c'í*  
 2SG(CL1) find:IPFV-2SG=ACC.CL2 child(CL2)  
 'You have found the child.'

- (33) Class 2 O, 3PL subject

*únī gām-án=ā dàwà kā rīs*  
 3PL(CL1) find:IPFV-3PL=ACC.CL2 baboon(CL2) with:CL2 many(CL2)  
 'They found many baboons.'

- (34) Class 2 O, 1SG person subject

*áhā p<sup>h</sup>ī-ná=ā sū*  
 1SG(CL1) drink:IPFV-1SG=ACC.CL2 beer(CL2)  
 'I am drinking the beer.'

Class 1 O arguments not only do not take overt Accusative marking, but they also trigger a reduction of verbal morphology. Subject cross-referencing markers on the verb for second and third person A arguments are suppressed,<sup>15</sup> and cross-referencing on the verb only appears with first person subjects.

- (35) Class 1 O, 3SG person subject

*ádĩ c'ít'-ø=ø bùnjè*  
 3SG(CL1) cut:IPFV-3SG=ACC.CL1 cloth(CL1)  
 'S/he's cutting the cloth.'

<sup>15</sup>Under normal circumstances, it is not possible for any other element to intervene between the verb and the noun that follows. There is one instance in my database pointed out to me by a reviewer (example 22), however, in which the aspectual marker *mò* does come in between a verb and a Class 1 noun. In this instance, cross-referencing of A on the verb is actually realized, suggesting that there may be additional factors involved in the suppression of the second/third person suffix. More research is needed to determine if this is indeed the case, and if so, what those might be. This may simply be an intransitive clause, with 'year' functioning adverbially.

- (36) Class 1 O, 3PL person subject

*únī d'ék'-∅=∅ k'wā*  
 3PL(CL1) pick\_up:IPFV-3PL=ACC.CL1 bowl(CL1)  
 'They pick up the bowl.'

- (37) Class 1 O, 2SG person subject

*é gǎm-∅=∅ tō yán*  
 2SG(CL1) find:IPFV-3SG=ACC.CL1 thing(CL1) DEM.PROX  
 'You found this thing.'

- (38) Class 1 O, 1SG person subject

*áhā p<sup>h</sup>ī-ná=∅ yid'é*  
 1SG(CL1) drink:IPFV-1SG=ACC.CL1 water(CL1)  
 'I am drinking the water.'

Examples (35), (36), and (37) are parallel to (31), (32), and (33) in structure, but with the subject cross-referencing markers on the verb suppressed.

First person subjects on the other hand do not change their cross-reference marking, irrespective of the gender of O. The only indication of the gender of O in examples (34) and (38) is the ACC marker.

The phenomenon described above does not apply to Narrative constructions, where arguments are never cross-referenced on the verb. This applies to all persons, with O arguments of either gender. Narrative constructions use non-finite forms of verbs, and the only difference between Narrative constructions with Class 1 objects and Narrative constructions with Class 2 objects is the Accusative case marker.

- (39) Class 1 O, Narrative construction

*à=cí kí k'ósh=∅ wàtí? mò*  
 CL2=creature(CL2) NARR hit<sub>NF</sub>=ACC.CL1 man(CL1) MO  
 'He attacks the man.'

- (40) Class 2 O, Narrative construction

*á'dī kí bùt=à c'í dālì k'ósh=ā c'í*  
 3SG(CL1) NARR catch<sub>NF</sub>=ACC.CL2 child(CL2) and hit<sub>NF</sub>=ACC.CL2 child(CL2)  
*mò*  
 MO  
 'She catches the child and beats the child.'

Note that personal pronouns have inherent Class 1 gender,<sup>16</sup> and the gender of a pronoun does not reflect the gender of the noun it denotes.

- (41) *à=k<sup>h</sup>úrā      lǒb=mā      c'í*  
 CL2=ball(CL2) play:IPFV=ERG.CL2 child(CL2)  
 'The child is playing football.'

- (42) *à=k<sup>h</sup>úrā      lǒb=ā      ádĩ*  
 CL2=ball(CL2) play:IPFV=ERG.CL1 3SG(CL1)  
 'S/he is playing football.'

Pronominal objects also trigger indexation patterns in which second and third person cross-referencing of A is suppressed.

- (43) Class 2 O, 3SG person subject  
*wàtĩ?      k'ǒf-ód=ā      Rǎbì*  
 man(CL1) hit:IPFV-3SG=ACC.CL2 Rabi(CL2)  
 'The man hits Rabi.'

- (44) Class 1 O, 3SG person subject  
*wàtĩ?      k'ǒf-ø=ø      ádĩ*  
 man(CL1) hit:IPFV-3SG=ACC.CL1 3SG(CL1)  
 'The man hits him/her/it.'

## 4 Gender assignment

Gender assignment in Uduk is largely, but not exclusively, arbitrary, with only limited connections to semantic categories such as biological sex, size, shape, and animacy. There are no distinctions based on sex, human vs. non-human, or animate vs. inanimate, and neither sex nor animacy is distinguished in the pronominal system for any person.

Nouns generally considered among the highest in the animacy scale, such as human kinship terms, do not show transparent assignment.

A list of human nouns and their gender may be found in Table 2, with little or no predictability beyond the fact that most suppletive possessive kinship terms appear to fall into Class 1.

---

<sup>16</sup>Described more fully in §4 below.

Table 2: Class 1 and Class 2 human nouns

Class 1	
<i>wàtí?</i>	man
<i>yà?</i>	son
<i>ḃwāhām</i>	female sibling or parallel cousin
<i>ḃwā?</i>	daughter
<i>āf</i>	wife
<i>jìl</i>	sisters-in-law, recip.
<i>kūm</i>	his, her mother
<i>kwān</i>	your mother
<i>cím</i>	your father
<i>cōm</i>	his, her father
<i>sób</i>	his, her father's sister
<i>nà(m)</i>	niece, nephew (sister's children)
<i>símín</i>	father's sister
<i>yàfím</i>	brother's wife; husband's brother or sister
<i>k'wáskām</i>	cross-cousin
<i>k'wáskīn</i>	your cross-cousin
all personal pronouns	
all plural derived agentive nouns	
Class 2	
<i>à=ḃóm</i>	woman, wife
<i>à=kām</i>	male sibling or parallel cousin
<i>à=bápá</i>	father
<i>à=tádā</i>	mother
<i>à=māmá</i>	my mother, also vocative
<i>à=kāṭ<sup>h</sup></i>	husband
<i>à=c'í</i>	child (general)
<i>à=māmà</i>	father's sister
<i>à=tāt<sup>h</sup>á</i>	mother's brother
<i>à=fwákām</i>	mother's brother
<i>à=nàrú</i>	mother's brother
<i>à=?íyā</i>	father's brother; brother's children
<i>à=màrè</i>	wife's parents
<i>à=màr</i>	husband's parents
<i>à=màsé?</i>	sister's husband
<i>à=m<sup>wí</sup></i>	sister's children (for men)
<i>à=dít<sup>h</sup>í?</i>	elderly woman, esp. father's sister
<i>à=mòpérù</i>	second cousin, more distant relationship
<i>à=pèrgòn</i>	cousin (non-vocative)
all personal names, male and female	
all singular derived agentive nouns	

Dahl (2000: 101) postulates the following:

1. In any gender system, there is a general semantically-based principle for assigning gender to animate nouns and noun phrases.
2. The domain of the principle referred to in (1) may be cut off at different points of the animacy hierarchy: between humans and animals, between higher and lower animals, or between animals and inanimates.

That is, by using a hierarchy such as the one found in Figure 1, one can make predictions on what types of gender systems may occur, and where semantically-based principles apply. Dahl suggests that cross-linguistic cut-off points vary, but are always found below human.

1st person	>	2nd person	>	3rd person	>	proper names	>	kin
		> other humans		> animate nouns		> inanimate nouns		

Figure 1: Animacy hierarchy

Semantic assignment is not predictable for human appellatives in Uduk; however, there *are* semantic areas in which predictability does occur: namely personal (and demonstrative) pronouns as well as proper names, both categories above human in the animacy hierarchy.

All personal pronouns show gender assignment in the same way that nouns do, and could be considered a lexical subtype of nouns. Demonstratives and personal pronouns are all assigned to the nominal Class 1 gender; they show no connection to the gender of a noun in anaphoric contexts, and are invariably Class 1. This is partially comparable to Jarawara (Arawan), in which “all pronouns (whatever the sex of their referent) engender feminine agreement on verbal suffixes” (Dixon 2000: 488). Proper names on the other hand are assigned to Class 2. This generalization holds only for personal names; place names can vary. Uduk gender predictability thus appears to apply only to levels higher than human appellatives in the animacy hierarchy.

Below this cut-off point there are limited trends in semantic assignment, but the semantic groups that can be formed all have exceptions. Nouns denoting plural entities, *k'wānī* ‘people’, *ūp<sup>h</sup>* ‘women’, and *ūc<sup>h</sup>i* ‘children’, are Class 1. Furthermore, a limited subset of nouns (primarily proper names and some kinship terms) in Uduk may appear with the Associative Plural prefix *ī-* to denote a person and additional people associated with that person; nouns marked in this way



are also Class 1. This includes plurals which would otherwise be assigned to Class 2, such as proper names.<sup>17</sup>

Most relational nouns, nouns which are primarily used to indicate more detailed types of spatial or temporal relationships, are also Class 1. This includes nouns like *fēmén* ‘alongside’, *p’émèn* ‘end, bottom (of)’, *bwàmán* ‘inside, between’, *bwàmbòr* ‘front (of)’; a few, such as *à=p’ó?* ‘on top of’ and *à=píjè* ‘outside’ are Class 2. Lastly, body parts are also more commonly found in Class 1 than Class 2.

Formal assignment in terms of word formation rules also creates limited situations in which gender assignment may be predicted. Nominalizations of stative verbs, marked with the suffix *-gá?*, are invariably assigned to Class 2. Agentive nouns formed with the derivational morpheme *màn-* are also assigned to Class 2. Nouns derived from verbs which use zero derivation, however, are all assigned to the Class 1 gender.

Uduk nouns tend to be fairly rigid in their assignment of gender, and few lexemes seem to have the possibility of occurring in either class. In these instances, there is no change in meaning. This includes intraspeaker variation as well as free variation within the speech of the same speakers.

There are a few instances in which homophonous nouns are assigned to different classes, e.g. *jè*, ‘elephant’, and *à=jè* ‘mud; type of fish’, but these are purely lexical distinctions, and remain rigid in assignment.

There is a markedness relationship between the two classes. In many respects, Class 1 could be considered the unmarked, default class, particularly for less prototypical nouns, such as pronouns. In addition to the lack of overt morphology in many instances, there are other signs that Class 1 is seen as the default. Conjunctions which occur before word classes other than nouns, for instance, use the same form as before Class 1 nouns. However, in other respects, Class 2 could also be considered a default. Class 2 is the default for nouns and adjective-like concepts, and a large number (although not all) of borrowed words appear to be placed into Class 2, e.g. *à=bāsál* ‘onion’, *à=bì/kír* ‘towel’, *à=màsábà* ‘distance’, *à=fábāgà* ‘network’.

## 5 Complexity

Uduk shows itself to have an atypical gender system, and it is worth investigating its complexity in more detail, and how it might compare to gender systems of other languages. Di Garbo (2014: 183) uses six features to determine the

<sup>17</sup>Note that most nouns in Uduk are not normally morphologically marked for number; the Associative Plural is one of very few ways of marking number directly on a noun, and even this is only possible to use with a limited set of nouns.

complexity of a gender system: Number of gender values, Nature of assignment rules, Number of targets, Cumulative exponence of gender and number, Manipulation of gender assignment triggered by number/countability, and Manipulation of gender assignment triggered by size.

In terms of these features as well as some others, Uduk has a relatively simple system. There are only two genders, to which nouns are generally rigidly assigned. No manipulation is possible, and aside from the Associative Plural marker, there are no instances in which number and gender are marked cumulatively. There are three targets: case marking particles, verbs, and adpositions/conjunctions/complementizers (which all form part of a single category), and a marginal fourth in the form of the diminutive (not included here as it does not constitute a word class; see §3.4). Assignment parameters feature higher complexity, however, as assignment is partly semantic, partly formal, but mostly completely opaque.

There were two additional criteria mentioned in §2, arbitrariness in gender assignment and adjacency, which play an interesting role in complexity, although at the moment it is difficult to see precisely how to reconcile them in terms of complexity metrics.

In nearly all instances in which gender is indexed on a target in Uduk, the gender-marked target and controller are immediately adjacent, with the target in the immediate position before the controller. This adds slightly to the descriptive complexity, as it requires an extra rule or constraint specifying this in the description.

Arbitrariness in gender assignment is even more difficult to reconcile, but an arbitrary system is likely also more complex. In principle, assignment would reach maximal complexity if each individual noun required a separate descriptive rule.

Both arbitrariness of assignment as well as adjacency require further research in general. Whether we exclude or include these as factors, however, it would appear that Uduk does have a relatively simple gender system, albeit atypical.

## **6 Discussion**

The Uduk gender system turns out to have a number of intriguing aspects. First, the system makes heavy use of zero marking and in one instance, suppression of subject agreement morphemes to indicate the gender of an object.

Second, almost all targets of indexation are adjacent to the controller. This is not commonly remarked upon cross-linguistically,<sup>18</sup> and by making note of it here, it may encourage other linguists to explore adjacency as a factor at play in gender marking systems.

Third, personal and demonstrative pronouns control gender in the same way that nouns do. And finally, gender is not connected to biological sex or other familiar semantic categories.

As mentioned previously, the last two characteristics are connected in Uduk. Semantic predictability in Uduk occurs at higher levels of animacy than simply human. It parallels some Austronesian languages such as Tagalog and Fijian for instance, which Hockett described as having gender, although later linguists have not.

In Fijian, /mata/ ‘day’ is preceded by /na/ when it is the subject of a clause, but /viti/ ‘Fiji’ is preceded instead by /ko/. /na/ and /ko/ are two distinct particles, not different inflected forms of a single stem. Yet the choice of /na/ or /ko/ establishes a twofold classification of all Fijian nouns and noun phrases: names of specific people and places belong to the /ko/ class, common nouns to the /na/ class. (Hockett 1958: 230)

Even more interestingly, “...independent pronouns [in Fijian] function in many ways like proper nouns, and are frequently marked by the same marker (*ko* or *o*)” (Geraghty 1983: 201).

A comparable system is found in Tagalog (Table 3), which could also be viewed as having a common vs. proper gender system. Tagalog additionally has distinct forms for demonstratives and each pronoun, suggesting that these are internally viewed as a third category, neither common nor proper (and different from Fijian in this respect).

In both cases, Tagalog and Fijian have a higher cut-off point in animacy than human nouns, requiring a more fine-grained approach to the animacy hierarchy. This cut-off point appears to show some parallels to Uduk. Where Fijian for instance differs from Uduk, however, is that in Uduk, proper names and personal pronouns do not occur in the same gender, and thus a proper-common gender differentiation would not be suitable as an analysis. Uduk would instead show two genders, one consisting of personal and demonstrative pronouns and other nouns, and the other consisting of proper names and other nouns.

---

<sup>18</sup>One important exception to this is Bernhard Wälchli’s work on Nalca (Wälchli 2018). Wälchli was also the one who pointed out adjacency as a relevant factor in Uduk to me, and I likely would not have noticed or remarked upon this without his input. Additionally, !Xóõ also appears to index gender only on adjacent targets; for further details, see Güldemann (2006).

Table 3: Noun phrase markers and pronouns in Tagalog (Himmelman 2005: 358)

	SPEC	POSS/GEN	LOC/DAT
Common nouns	<i>ang</i>	<i>ng</i>	<i>sa</i>
Personal names	<i>si</i>	<i>ni</i>	<i>kay</i>
1SG	<i>akó</i>	<i>ko</i>	<i>akin</i>
2SG	<i>ikáw, ka</i>	<i>mo</i>	<i>iyó, iyó</i>
3SG	<i>siyá</i>	<i>iyá</i>	<i>kaniyá</i>
1DU.IN	<i>kitá, katá</i>	<i>nitá</i>	<i>kanitá</i>
1PL.IN	<i>tayo</i>	<i>natin</i>	<i>atin</i>
1PL.EX	<i>kamí</i>	<i>namin</i>	<i>amin</i>
2PL	<i>kayó</i>	<i>ninyó</i>	<i>inyó</i>
3PL	<i>silá</i>	<i>nilá</i>	<i>kanilá</i>
PROX	<i>itó</i>	<i>nitó</i>	<i>dito, rito</i>
MED	<i>iyán</i>	<i>niyán</i>	<i>diyán, riyán</i>
DIST	<i>iyón</i>	<i>niyón, noón</i>	<i>doón, roón</i>

Languages like Tagalog, Fijian, and Uduk give evidence suggesting that predictability may occur at points higher in the animacy hierarchy than previously acknowledged, although Uduk shows itself to be more complex than Tagalog or Fijian, as the gender of its nouns is generally much less predictable. By including Uduk as a typological point of reference, a reconsideration of possible cut-off points in the animacy hierarchy may be in order.

## Special abbreviations

The following abbreviations are not found in the Leipzig Glossing Rules:

AD1	aspect-directional 1	MO	aspect-mood particle
AD2	aspect-directional 2	NARR	narrative
ASSOC	associative	NAS	nasal
CF	counterfactual	NF	non-finite
CL1	class 1 gender	PART	partargument
CL2	class 2 gender	PP	predicative possession
DIM	diminutive	REL	relativizer
LNK	linker	SPEC	specific article

## Acknowledgements

I would like to thank the Kone foundation, who financially supported this research, and the UH-SU cooperation project, who financially supported the workshop on Grammatical Gender and Linguistic Complexity. I would also like to thank Bernhard Wälchli and Francesca Di Garbo for all their time and effort as both reviewers as well as editors; they contributed a great deal in helping me develop the ideas presented here. I also thank Bruno Olsson for his help and technical assistance in editing and layout formatting, and Johanna Nichols, Manuel Otero, the anonymous reviewers, and the participants of the Grammatical Gender and Linguistic Complexity workshop for their feedback and comments. Last but not least, I would also like to thank all of my Uduk consultants, who devoted a lot of time and effort to helping me understand their language.

Any remaining errors are of course the author's own responsibility.

## References

- Audring, Jenny. 2019. Canonical, complex, complicated? In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 15–52. Berlin: Language Science Press. DOI:??
- Beam, Mary S. & Elizabeth A. Cridland. 1970. *Uduk-English dictionary*. Khartoum: Sudan Research Unit, University of Khartoum.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Cysouw, Michael. 2005. Morphology in the wrong place: A survey of preposed enclitics. In Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer & Franz Rainer (eds.), *Morphology and its demarcations: Selected papers from the 11th Morphology Meeting, Vienna, February 2004*, 17–37. Amsterdam: John Benjamins.
- Dahl, Östen. 2000. Animacy and the notion of semantic gender. In Barbara Unterbeck (ed.), *Gender in grammar and cognition*. Vol. 1: *Animacy and the notion of semantic gender: Approaches to gender*, 99–115. Berlin: Mouton de Gruyter.
- Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
- Di Garbo, Francesca. 2014. *Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of Africa*. Stockholm University dissertation.
- Di Garbo, Francesca. 2016. Exploring grammatical complexity crosslinguistically: The case of gender. *Linguistic Discovery* 14(1). 46–85.

- Dixon, Robert M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press.
- Dixon, Robert M. W. 2000. Categories of the noun phrase in Jarawara. *Journal of Linguistics* 36(3). 487–510.
- Geraghty, Paul. 1983. *History of the Fijian languages* (Oceanic Linguistics Special Publications 19). Honolulu: University of Hawai'i Press.
- Greenberg, Joseph H. 1963. *The languages of Africa*. Bloomington: Indiana University.
- Güldemann, Tom. 2006. "Janus-headed" nominals: The morphosyntax of agreement in Taa. Paper presented at the International Conference "Rara & Rarissima – collecting and interpreting unusual characteristics of human languages", Leipzig, 29 March–1 April 2006.
- Himmelman, Nikolaus. 2005. Tagalog. In Alexander Adelaar & Nikolaus Himmelman (eds.), *The Austronesian languages of Asia and Madagascar*, 350–376. New York: Routledge.
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.
- James, Wendy. 1979. 'Kwanim pa. – *The making of the Uduk people: An ethnographic study of survival in the Sudan-Ethiopian borderlands*. Oxford: Clarendon Press.
- Killian, Don. 2015. *Topics in Uduk phonology and Morphosyntax*. University of Helsinki dissertation.
- 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.
- Wälchli, Bernhard. 2018. The rise of gender in Nalca (Mek, Tanah Papua): The drift towards the canonical gender attractor. In Sebastian Fedden, Jenny Audring & Greville G. Corbett (eds.), *Non-canonical gender systems*, 68–99. Oxford: Oxford University Press.

## **Part III**

# **New Guinea**





## Chapter 7

# Gender in Walman

Matthew S. Dryer

University at Buffalo

In this paper, I describe gender and gender-like phenomena in Walman, a language of the Torricelli family spoken on the north coast of Papua New Guinea. I discuss three topics. One of these is the two clear instances of gender in Walman, masculine and feminine. I discuss the formal realization of gender in Walman and the factors governing the choice of masculine versus feminine gender.

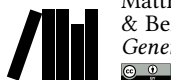
There are also two gender-like phenomena in Walman, namely pluralia tantum nouns and a diminutive category. Pluralia tantum nouns in Walman are different from pluralia tantum nouns in European languages in that what makes them grammatically plural is not their form, but the fact that they control plural agreement. What makes pluralia tantum gender-like is that there are twice as many pluralia tantum nouns in our data as there are nouns that are lexically masculine.

The second gender-like phenomenon in Walman is a diminutive category, which is coded in the same way as feminine singular, masculine singular, and plural. What makes it unlike phenomena that are normally considered instances of gender in other languages is the fact that there are no lexically diminutive nouns and any noun can be associated with diminutive agreement.

**Keywords:** gender, masculine, feminine, diminutive, pluralia tantum, Walman, Torricelli.

## 1 Introduction

The goal of this paper is to give a description of gender in Walman, a language in the Torricelli family spoken in Papua New Guinea. I understand gender to denote a morphosyntactic category in a language based on a division among nouns in the language and on agreement phenomena related to this division. There are two unambiguous instances of genders in Walman, namely masculine and feminine. But there are also two other gender-like phenomena in the language, namely



pluralia tantum nouns and a diminutive category. I will describe the first of these phenomena in some detail in this paper, discussing ways in which it is like or unlike clear instances of gender. My discussion of the diminutive category will be briefer, since it is discussed in more detail elsewhere, in Dryer (2016) and Dryer (under revision).

In §2, I provide a brief grammatical sketch, primarily describing inflectional categories that vary for gender. In §3, I describe the factors governing the choice between masculine and feminine gender. In §4, I describe pluralia tantum nouns in Walman and in §5, I briefly describe the Walman diminutive.

## 2 Brief grammatical sketch

This section focuses primarily on the coding of gender in Walman, along with the coding of number, person, and diminutiveness. See Dryer (n.d.) for a description of other features of Walman.

Verbs in Walman inflect for both subject and object (and in some applicative constructions, for two objects). The subject affixes are word-initial prefixes consisting of single consonants, as in (1), where the verb *mara* ‘come’ bears a 1SG subject prefix *m-* and the verb *nawa* ‘call’ bears a 2SG subject prefix *n-*.

- (1) *Kum m-ara            eni        chi n-awa.*  
 1SG 1SG.SUBJ-come because 2SG 2SG.SUBJ-call  
 ‘I came because you called.’

Example (2) contains two occurrences of the 1PL subject prefix *k-*.

- (2) *Akou k-anan        k-ara        komoru.*  
 finish 1PL-go.down 1PL-come evening  
 ‘Then we walked home in the afternoon.’

The 2PL subject prefix *ch-* is illustrated in (3).<sup>1</sup>

- (3) *Chim ch-orou nyien?*  
 2PL 2PL-go where  
 ‘Where are you (plural) going?’

Example (4) contains two occurrences of the 3PL subject prefix *y-*.

---

<sup>1</sup>Our orthography for Walman employs three digraphs, <ch> for [tʃ], <ng> for [ŋ], and <ny> for [ɲ].

- (4) *Ri pelen y-anan y-okorue wul.*  
 3PL dog 3PL-go.down 3PL-bathe water  
 ‘Then the dogs went in for a wash.’

As mentioned above, there are two clear cases of gender in Walman, masculine and feminine; this distinction is realized only in the 3SG. Example (5) illustrates the 3SG.M subject prefix *n-* (again occurring twice).

- (5) *Runon n-rukuel n-anan nyuey.*  
 3SG.M 3SG.M-run 3SG.M-go.down sea  
 ‘He ran to the beach.’

And (6) illustrates the 3SG.F subject prefix *w-*.

- (6) *Nakol kkuk w-anan.*  
 house broken 3SG.F-go.down  
 ‘The house fell down.’

There is also a diminutive subject prefix *l-*, illustrated on *lakor* ‘drown’ in (7).

- (7) *Nyanam mon ro-l, ampa rul l-akor wul.*  
 child NEG tall-DIMIN FUT 3.DIMIN 3.DIMIN-drown water  
 ‘The child is small, she will drown.’

Although the diminutive is like masculine and feminine in being restricted to singular, it involves a distinct notion of ‘singular’, as discussed in §5 below.

There is also a set of object affixes that occur on transitive verbs, though they occur in three different positions within the verb. The first and second person object affixes are prefixes that immediately follow the subject prefixes. These prefixes are unspecified for number and are illustrated in (8) by the first person object prefix *p-* and in (9) by the second person object prefix *ch-*.

- (8) *Kum m-alma, chim ch-p-chami.*  
 1SG 1SG-die 2PL 2PL-1OBJ-bury  
 ‘When I die, bury me.’
- (9) *Opucha mol w-ch-any chi?*  
 thing which 3SG.F-2OBJ-happen.to 2SG  
 ‘What happened to you?’

A reflexive/reciprocal prefix /r/ occurs in the same slot as the first and second person object prefixes, as illustrated by the verb *yrklwaro* ‘they deceived each other’ in (10).

- (10) *Kamte-n ngo-n w-ri Walis n-aro-n nyemi kasim*  
 person-M one-M GEN-3PL Walis 3SG.M-and-3SG.M friend friend  
*y-r-klwaro.*  
 3PL-REFL/RECIP-deceive  
 ‘A man from Walis Island and his friend deceived each other.’

The third person object affixes are generally suffixes, though with a minority of verbs they are infixes. Examples (11) and (12) illustrate the 3PL and 3SG.M object suffixes respectively.

- (11) *Kum m-ete-y wuel chomchom.*  
 1SG 1SG-see-3PL.OBJ pig many  
 ‘I saw many pigs.’

- (12) *Ru w-lro-n runon.*  
 3SG.F 3SG.F-like-3SG.M:OBJ 3SG.M  
 ‘She likes him.’

The form of the third person object affixes is, with one exception, the same as the corresponding subject prefixes. For example, /n/ is the form of both the 3SG.M subject prefix, as in (5) above, and the 3SG.M object affix, as in (12). The one difference between the third person subject prefixes and third person object affixes is in 3SG.F, where the subject prefix is *w-*, as in (12), while the object affix is phonologically null, as illustrated by the form *mete* ‘see’ in (13) (contrasting, for example, with the presence of an overt object suffix for 3PL in the form *metey* in (11)).

- (13) *Kum m-ete-Ø chuto nyanam.*  
 1SG 1SG-see-3SG.F woman child  
 ‘I saw a young girl.’

With some verbs, the third person object affixes are infixes, as in the form *yanpu* ‘kill’ in (14), where the 3SG.M object affix *-n-* is an infix inside the verb stem *-apu* ‘kill’.

- (14) *Rim y-a<n>pu          ampatu          mon nngkal.*  
 3PL 3PL-kill<3SG.M> ground.wallaby NEG small  
 ‘They killed a big wallaby.’

Inflection for gender, as well as number and diminutiveness, also occurs on some adnominal words, including a small subset of adjectives, a subset of demonstratives and two numeral words meaning ‘one’.<sup>2</sup> The form of affixes indicating gender, number, or diminutiveness on adnominal words is the same as those used for object affixes on verbs. In (15), for example, we find the masculine affix *-n-* as an infix in the demonstrative *panten* and as a suffix on the adjective *lapo* ‘big’ (here used predicatively).

- (15) *Ngolu          pa<n>ten n-o          lapo-n.*  
 cassowary that<M> 3SG.M-be big-M  
 ‘That cassowary is large.’

Like the 3SG.F object affix on verbs, feminine gender is phonologically null on adnominal words, as illustrated by the feminine forms *paten* ‘that’ in (16) and *lapo* ‘big’ in (17).

- (16) *Mon chi n-a<ø>ko          wul          pa<ø>ten.*  
 NEG 2SG 2SG-eat<3SG.F> water that<F>  
 ‘You shouldn’t drink that water.’
- (17) *Wako lapo-ø w-ara.*  
 boat large-F 3SG-come  
 ‘A big ship has come.’

In (18), we get a plural suffix *-y* on *lapoy* ‘good’.

<sup>2</sup>There are five adjectives that inflect for gender: *lapo* ‘large’, *nyopu* ‘good’, *woyue* ‘bad’, *wwe* ‘bad’, and *kolue* ‘short’. The meanings associated with these correspond closely to the adjectival concepts found in languages with small adjective inventories (Dixon 1977). One might expect to find adjectives meaning ‘small’ or ‘long’ in this set. The Walman adjective for ‘small’, *nngkal*, does not inflect for gender but does for number; the plural form is *nngkam*. The meaning of ‘long, tall’ in Walman is expressed by a sequence of two words *ro rani*, where *ro* exists separately as an adnominal word meaning ‘piece of’ and does inflect for gender, so the two word sequence (feminine *ro rani*, masculine *ron rani*) can be described as functioning as an adjective and hence as a sixth adjective that inflects for gender.

- (18) *Nypeykil lapo-y y-an olun olun.*  
 tree.PL big-PL 3PL-be.at side side  
 ‘There are big trees on both sides of the road.’

There is no gender distinction in the plural. Note that the position of these affixes is similar to the position of corresponding object affixes in being typically suffixes (as in *lapon* in (15) and *nyopuy* in (16)), but with some words infixes (as in *panten* in (15)).

There are also two words for ‘one’ that inflect for gender, number, and diminutiveness, illustrated by *alpan* ‘one’ in (19).

- (19) *Kamte-n alpa-n n-epin n-ara.*  
 person-M one-M 3SG.M-go.ahead 3SG.M-come  
 ‘One man came ahead of the others.’

Not all adnominal words inflect. In fact most adjectives do not. For example the adjective *chapa* ‘fat’ is invariant, as illustrated in (20) (where the form would be a masculine form *chapan* if it did inflect).

- (20) *Runon n-o chapa.*  
 3SG.M 3SG.M-be fat  
 ‘He is fat.’

Finally, the third person pronouns themselves vary for number, gender, and diminutiveness, as illustrated by the pronouns for 3SG.M, *runon*, in (20) and 3SG.F, *ru*, in (12) above.

The only morphology found on nouns is plural marking.<sup>3</sup> However, plural marking occurs with a relatively small number of nouns; most nouns lack distinct plural forms. The set of nouns with distinct plural forms includes most kinship terms and a few other nouns denoting humans, plus seventeen inanimate nouns. There seems little way to predict which inanimate nouns have distinct plural forms. Some are nouns denoting body parts (e.g. *kampotu* ‘knee’, plural *kamtikiel*). Others include *nyikie* ‘piece of wood’, plural *nyikiel*; *nymuto* ‘star’,

<sup>3</sup>There are a few words that might be analysed as nouns that inflect for gender, since they involve a contrast that is formally identical to gender inflection on many adnominal words. First, there is a noun *kamten* ‘man’ with plural *kamtey* for which we have a few instances of a feminine form *kamte* and a diminutive form *kamtel* in elicited data, but none in texts. Second, there are a few pairs of kin terms differing in that the one denoting a male ends in an /n/ while the corresponding one denoting a female lacks the /n/, like *wlapon* ‘older brother of a man’ and *wlapo* ‘older sister of a woman’.

plural *nymteykil*; and *tomuel* ‘stone’, plural *tmleykiel*. The process of plural formation is fairly irregular. There are no plural forms for nouns denoting non-human animals. Whether a noun has a distinct plural form or not has no effect on agreement patterns. For nouns lacking distinct plural forms, differences in number are carried only on agreeing words. For example, what conveys the difference in number in (21) and (22) is the subject prefix on the verb (*w-* for 3SG feminine in (21), *y-* for 3PL in (22)); the form of the noun *pelen* ‘dog’ is the same in the two examples.

- (21) *Pelen w-aykiri.*  
 dog 3SG.F-bark  
 ‘The dog (female) is barking.’

- (22) *Pelen y-aykiri.*  
 dog 3PL-bark  
 ‘The dogs are barking.’

Among other grammatical features of Walman illustrated by the above examples is the fact that the language lacks case marking to distinguish arguments in a clause and the fact that the most frequent word order is SVO (though SOV exists as a not uncommon alternative order). Apart from the subject and object affixes described above, the only other verb morphology is an applicative suffix and a largely obsolete imperative form of verbs.

### 3 Principles of gender assignment

In (23) is a summary of the principles governing the choice between masculine and feminine gender in Walman.

- (23) a. All nouns denoting humans and some larger animals are either masculine or feminine, depending on the sex of the referent  
 b. All nouns denoting inanimate objects are feminine<sup>4</sup>  
 c. Nouns denoting a few quasi-animate natural phenomena, such as *nganu* ‘sun’, are masculine  
 d. Nouns denoting most animals appear to have relatively arbitrary gender

---

<sup>4</sup>As discussed below in §4, there are many nouns denoting inanimate objects which are pluralia tantum nouns. These nouns are neither masculine nor feminine.

The first principle, given in (23a), is that all nouns denoting humans and some larger animals can be either masculine or feminine, depending on the sex of the referent.<sup>5</sup> For example, the noun *pelen* ‘dog’ controls feminine subject agreement in (24), but masculine subject agreement in (25).

- (24) *O pelen tu w-ata ke?*  
 and dog PERF 3SG.F-bite.2OBJ Q  
 ‘Did the dog bite you?’

- (25) *Kum wuel mingrieny tu pelen n-a<y>ko.*  
 1SG pig meat PERF dog 3SG.M-eat<3PL>  
 ‘My pig’s meat has been eaten by the dog.’

Most nouns denoting humans are inherently masculine or feminine, but only because they necessarily denote someone who is male or female respectively. For example, in (26), the noun *ngan* ‘father’ controls masculine subject agreement on *nroko* ‘take’ while *nyue* ‘mother’ controls feminine subject on *wrulu* ‘cut’.

- (26) *Ngan n-r-oko rele, nyue w-r-ulo woruen.*  
 father 3SG.M-REFL-take beard mother 3SG.F-REFL-cut hair  
 ‘The father shaves, the mother trims her hair.’

The second principle is that nouns denoting inanimate objects are feminine. This is illustrated in (27), where *chakonu* ‘road’ controls 3SG.F agreement on the verb *wo* ‘be’.

- (27) *Chakonu w-o mail.*  
 road 3SG.F-be crooked  
 ‘The road is not straight.’

This principle is also illustrated in examples above, for *nakol* ‘house’ in (6), for *opucha* ‘thing’ in (9), and for *wul* ‘water’ in (16).

What could be interpreted as an exception to this principle is stated above in (23c): nouns denoting a few quasi-animate natural phenomena are masculine.

<sup>5</sup>The only nouns denoting animals for which we have clear evidence on this are the nouns *pelen* ‘dog’ and *wuel* ‘pig’. There are some other nouns, like *slaoi* ‘rat’, where some instances in our data control masculine agreement and others control feminine agreement, but we need to investigate to determine whether this alternation is governed by the presumed sex of the referent (or some other factors).



This is illustrated for *snar* ‘moon’ in (28), where it controls masculine subject agreement, and for *onyul* ‘earthquake’ in (29), where it controls masculine object agreement.

- (28) *Snar n-reliel.*  
 moon 3SG.M-shine  
 ‘The moon is shining.’

- (29) *Kum m-rere-n onyul nngkal.*  
 1SG 1SG-feel-3SG.M earthquake small  
 ‘I felt a small earthquake.’

There are two other nouns of this sort that consistently control masculine agreement, namely *nganu* ‘sun’ and *knum* ‘whirlpool, riptide’. Note that *nganu* ‘sun’ can also mean simply ‘day’ and controls masculine agreement with this meaning as well, as in (30), where it controls masculine agreement on the adnominal word *ngon* ‘one’, as reflected by the masculine suffix *-n*.<sup>6</sup>

- (30) *Nganu ngo-n ru w-ekete-n chamul w-ru.*  
 sun one-M 3SG.F 3SG.F-pull-3SG.M Chamul GEN-3SG.F  
 ‘One day she played a flute to call her Chamul.’

There are two other nouns of this sort that can control masculine agreement, but only when they occur in idioms, not when they occur with their literal meaning. One is the noun *olokol* ‘mountain’, which is normally a pluralia tantum noun, controlling plural agreement, as in (31), where it controls plural inflection on *alpay* ‘one’ and 3PL subject agreement on the verb *yiliel* ‘go towards sea’.<sup>7</sup>

- (31) ... *olokol alpa-y konu y-iliel Matapau.*  
 mountain one-PL only 3PL-go.seaward Matapau  
 ‘... there was just one mountain coming down at Matapau.’

However, this noun also occurs with the verb *-oruel* ‘explode’ in an idiom meaning ‘to thunder’, as in (32), and in this idiom it controls masculine subject agreement on the verb.

<sup>6</sup>A *chamul* is a partly human, partly supernatural being in traditional Walman culture. Example (30) employs an idiom *-ekete chamul* ‘to play a flute to call one’s *chamul*’.

<sup>7</sup>Normally *olokol* refers to an entire mountain range, since the salient mountains near Walman-speaking villages are the Torricelli Mountains, a mountain range that is roughly parallel to the coast, where there is not a clear delineation between individual mountains. In (31), however, it is clear from the text that this comes from that a single mountain is being referred to.

- (32) *Olokol n-oruel.*  
 mountain 3SG.M-explode  
 ‘It thundered.’

In other contexts with the verb *-oruel*, this noun triggers plural subject agreement, but in these cases, the meaning is literal rather than idiomatic, as illustrated in (33).

- (33) *Olokol y-oruel.*  
 mountain 3PL-explode  
 ‘The mountain exploded (i.e. a volcano).’

The second noun that controls masculine agreement in an idiom but not in its literal meaning is the noun *anako* ‘sky’, which combines either with the verb *-ol* ‘break’ or with the verb *ochoro* ‘split open’ as alternative ways to express the meaning ‘to thunder’, as illustrated with the verb *-ol* in (34).

- (34) *Anako n-ol komoru.*  
 sky 3SG.M-break evening  
 ‘It thundered in the (late) afternoon.’

Outside of this idiom, the noun *anako* ‘sky’ controls feminine agreement, as illustrated in (35).

- (35) *Lasi anako w-arau w-orou wor.*  
 immediately sky 3SG.F-go.up 3SG.F-go high  
 ‘The sky immediately went high up.’

Although these nouns denote things that are considered inanimate in Western cultures, I characterize them as quasi-animate, since they all denote things that are associated with autonomous movement or force, something generally associated with animate beings. However, not all nouns that might be considered instances of autonomous movement or force control masculine agreement, as illustrated for *loun* ‘cloud’ in (36) and for *nyuey* ‘sea’ in (37), which are both feminine, as reflected by the 3SG subject prefixes *w-* on the verbs.

- (36) *Loun w-alplo-n nganu.*  
 cloud 3SG.F-cover-3SG.M sun  
 ‘The cloud is hiding the sun.’

- (37) *Nyuey w-oko-n n-orou w-elie-n n-ekiel ...*  
 sea 3SG.F-take-3SG.M 3SG.M-go 3SG.F-throw-3SG.M 3SG.M-go.landward  
 ‘The sea carried him until it threw him up on the beach ...’

Another noun, *chepili* ‘thunder, lightning’, always controls plural agreement, as in (38), where it controls 3PL subject agreement on *yol* ‘break’, *yanan* ‘go down’ and *yaypu* ‘kill’.<sup>8</sup>

- (38) *Ru w-ao-y nyiki, lasi chepili y-ol*  
 3SG.F 3SG.F-shoot-3PL woman.PL immediately thunder 3PL-break  
*mpang, y-anan, y-a<y>pu kamte-y eni y-a<ø>ko*  
 loud.noise 3PL-go.down 3PL-kill<3PL> person-PL REL 3PL-eat<3SG.F>  
*wkaray w-aro-ø ngotu, y-alma mpor.*  
 white.cuscus 3SG.F-and-3SG.F coconut 3PL-die all  
 ‘There was lightning and immediately thunder cracked “mpang” and came down and killed all the people who had eaten the cuscus with coconut.’

The only nouns in Walman for which gender appears to be arbitrarily assigned are those denoting other animals, especially non-mammals. For example, *alan* ‘red and green parrot’ is masculine, as reflected by the masculine subject prefixes on the verbs *nka* ‘fly’ and *nekiel* ‘go inland, go towards land’ in (39).

- (39) *Alan yapa n-ka n-ekiel.*  
 parrot that 3SG.M-fly 3SG.M-go.landward  
 ‘That parrot is flying inland.’

Similarly *wraul* ‘toad’ is feminine, as reflected in (40) by the feminine object agreement on *nete* ‘see’, the feminine agreement on the adjective *lapo* ‘large’, and the feminine subject agreement on *wekele* ‘make’.

- (40) *Lasi runon Tenten n-ete-ø wraul lapo-ø oluel*  
 immediately 3SG.M Tenten 3SG.M-see-3SG.F toad big-F nest  
*w-ekele w-an kra nyumuen.*  
 3SG.F-make 3SG.F-be.at sugarcane middle  
 ‘A man Tenten suddenly saw a large toad making a nest in the middle of the sugarcane.’

<sup>8</sup>The first three words in (38) constitute an idiom meaning ‘for there to be lightning’, where the literal meaning is ‘it shoots women’. Note that this idiom obligatorily has the 3SG.F pronoun *ru* as subject.

For a number of reasons, it is not really possible to demonstrate convincingly that gender is arbitrary for most animals. First, for many species, we have not actually seen instances of the animals, but depend on descriptions by speakers. Second, one can never know for sure whether there are unknown characteristics of particular animals that play a role in determination of gender (such as size, sound, or behaviour). And third, there may be roles that animals play in Walman culture and history that we are not aware of that influence gender. In general, however, native speakers do not have explanations for particular gender assignment for these nouns.

The lack of an obvious semantic basis for gender assignment for animals can be illustrated by looking at the gender of nouns denoting various species of snakes. In (41), I list the genders for the six nouns (or two-word nominal expressions) in our data denoting different species of snake.

(41) Snakes

MASCULINE

- |                   |  |
|-------------------|--|
| <i>ani konu</i>   | snake, light brown-orange-red, about a metre long, very dangerous                              |
| <i>nayko iyoy</i> | small snake, lives along coast, not dangerous, eats crabs                                      |
| <i>layat</i>      | type of python, very big and long, pretty patterned skin, lives in trees, not really dangerous |

FEMININE

- |                 |  |
|-----------------|--|
| <i>kilekile</i> | death adder, about a foot long, black with white dots, very dangerous                                      |
| <i>mekey</i>    | ground python, brown with white belly, not poisonous, can be very big                                      |
| <i>nyieu</i>    | very big and long, light blue and shiny, lives in bush, not dangerous to people but swallows small animals |

Two obvious differences among snakes that might play a role in determining gender are size and how dangerous they are (defined by how serious their snake bite is). The list of snakes in (41) includes three pythons, which share the features of being large and not being dangerous: two are masculine, while one is feminine. Of the three smaller snakes, two are very dangerous: one of these is masculine, the other feminine. Thus neither size nor how dangerous they are provides a basis for predicting gender. There may be other factors, of course, but the most obvious ones do not seem relevant. Note that *ani konu* is literally ‘male snake’, so the masculine gender for this two-word nominal expression is explained by

the fact that *konu* means ‘male’. In addition the first word in *nayko iyoy* is a form that looks like a form of the verb *-ako* ‘eat’, with a 3SG.M prefix and a 3PL object infix, while the second word (*iyoy*) is a noun meaning ‘crab’ so that the apparent literal meaning of *nayko iyoy* is ‘he eats crabs’; thus the fact that *nayko* begins with what looks like a 3SG.M subject prefix may be relevant to the fact that this snake is masculine.

We find a similar situation with insects and similar lower animals. The list in (42) is a list of all the species of such animals in our data (excluding a few whose gender we lack data on).

(42) Insects and the like (spiders, lice, leeches, worms, centipedes, millipedes)

MASCULINE

<i>achakol</i>	housefly
<i>kayikiel</i>	fruitfly
<i>kaimung</i>	firefly
<i>kanal</i>	sago grub
<i>melkil</i>	bee, wasp
<i>mile</i>	leech
<i>paral tkay</i>	flying ant
<i>ppu</i>	small green or brown grasshopper-like creature
<i>slmako</i>	bluebottle fly
<i>srnyako</i>	beetle which comes around in evening, makes loud sound
<i>tmpinie</i>	worm (general term)

FEMININE

<i>atal</i>	scorpion
<i>inrer</i>	very small mosquito (hard to see, smaller than sandfly) that bites people in evening, especially in marshy areas of bush
<i>klu</i>	‘fly which is very tiny and which makes its nest in holes in wood’
<i>krunu</i>	centipede
<i>nymuchuto</i>	spider
<i>nymulol</i>	louse
<i>paral</i>	ant
<i>pirinyue</i>	cockroach

<i>posur</i>	termite, white ant (does not live in houses, builds mounds)
<i>puseksek</i>	a type of grasshopper that is large, brown or green, and can fly, and which makes a noise like “seksek”
<i>waykelie</i>	millipede
<i>woru</i>	mosquito

The nouns listed in (42) denoting species which bite or sting humans include three masculine nouns (*melkil* ‘bee, wasp’, *mile* ‘leech’, and *paral tkay* ‘flying ant’) and seven feminine nouns (*atal* ‘scorpion’, *inrer* ‘very small mosquito’, *krunu* ‘centipede’, *nymuchuto* ‘spider’, *nymulol* ‘louse’, *paral* ‘ant’, and *woru* ‘mosquito’), so being something that bites or stings is not a predictor of gender. Of the two species whose stings are most painful, one is masculine (*melkil* ‘bee, wasp’) while the other is feminine (*krunu* ‘centipede’, the local variety of which reportedly has an especially painful sting). Of the smaller species in (42), one is masculine (*kayikiel* ‘fruit fly’) while three are feminine (*inrer* ‘very small mosquito’, *klu* ‘very tiny fly’, and *woru* ‘mosquito’). Nor is there any other obvious feature distinguishing the masculine nouns in (42) from the feminine nouns.

If there is any feature that correlates at least weakly with gender among other animals, it is that nouns denoting more aggressive species are somewhat more often masculine while nouns denoting less aggressive species are somewhat more often feminine. A correlation with aggressiveness seems most apparent with species of birds, listed in (43).

(43) Birds

MASCULINE

<i>alan</i>	red and green parrot
<i>aron</i>	eagle that is large and grey and white and that is found in the jungle
<i>mmpul</i>	hawk with reddish brown body and white head
<i>ngolu</i>	cassowary
<i>semier</i>	type of bush fowl
<i>tarkau</i>	osprey
<i>tualiau</i>	type of bush fowl, brown, small, the size of a chicken
<i>wamol</i>	hornbill
<i>wawiel</i>	crow
<i>yiwos</i>	very small hawk, brown, lives at coast

## FEMININE

<i>kmaynum</i>	blue bird about the size of a chicken, has no decoration
<i>le</i>	bird of paradise
<i>pinie</i>	tiny bird, blue with white around neck
<i>polmonu</i>	guria pigeon
<i>rampanyau</i>	willy wagtail
<i>solponyou</i>	swallow
<i>yup</i>	white cockatoo

All of the nouns denoting what I believe are the most aggressive species are masculine: *nganu* ('cassowary'), *aron* (a type of eagle), *mmpul* (a type of hawk), *yiwos* (another type of hawk), *tarkau* ('osprey'), and *wawiel* ('crow').

Most of the nouns denoting aquatic animals are feminine. This includes nine out of twelve species of fish, two species of crab, crayfish, and two aquatic mammals (*alpariak* 'dolphin', *yuel* 'seal'). One of the three masculine nouns for a species of fish is the noun *wuey* for 'shark', which fits the weak correlation between aggressiveness and masculine gender. There is one noun, *nyekekēl*, that can denote either of two species of snail. This noun is masculine when it denotes one species, feminine when it denotes the other species. The feminine one lives in water, while the masculine one apparently does not.

Some nouns denoting larger animals can be either masculine or feminine, but one of the two genders is the default. While it is apparently the case that the default gender is generally used when the sex of the referent is unknown, this is not always the case. For example the default gender of the noun *ngolu* 'cassowary' is masculine and although it can be feminine when the referent is female, feminine gender is not obligatory when the referent is clearly female. In (44), for example, this noun controls masculine subject agreement on the verb, despite the fact that the semantics of the sentence implies that the referent is female.

- (44) *Ngolu*      *n-ikie-ø*              *meten*.  
 cassowary 3SG.M-put-3SG.F egg  
 'A cassowary has laid an egg.'

However, this noun can be feminine, as in (45), where it controls feminine object agreement.<sup>9</sup>

<sup>9</sup>The possibility of feminine agreement in (45) may be due to the fact that it is the meat (i.e. an inanimate object) that is being denoted here, rather than the living bird. However we have more than one other instance in our data of a noun phrase denoting cassowary meat triggering masculine agreement.

- (45) ... *y-e<ø>tiki* *ngolu.*  
 3PL-cook.over.fire<3SG.F> cassowary  
 '[She is still with her brothers] cooking (a) cassowary.'

There are a few other uses of masculine gender in Walman that are more unusual. For example, the noun *won* can mean 'chest', but it is far more common as part of a large number of idioms where this meaning is less evident. In its meaning 'chest', it is feminine, as in (46).

- (46) *Won mnon w-o lapo-ø.*  
 chest 3SG.M:GEN 3SG.F-be big-F  
 'His chest is large.'

When *won* occurs in idioms, it is masculine, as in (47) and (48), where in both cases *won* controls masculine subject agreement on the verb. The idiom in (47) for 'angry' is literally 'heart be fast'.

- (47) *Ru won n-o kisiel prie.*  
 3SG.F heart 3SG.M-be fast completely  
 'She is very angry.'

I gloss *won* in idioms as 'heart', not in the sense of the body part, but in a more abstract sense that could alternatively be glossed 'mind' or 'soul'. One reason that I gloss it as 'heart' is that it is clearly cognate to the word for the body part heart in a number of other languages in the Torricelli family.

The idiom in (48) for 'be happy' is literally 'heart follows', where the one who is happy is grammatically the object of the verb, as reflected by the 3PL object suffix on the verb. Note that the object pronoun *ri* in (48) is clause-initial; the normal word order in this and a couple of other idiomatic constructions with an inanimate subject and an animate object is OSV.

- (48) *Ri won n-rowlo-y.*  
 3PL heart 3SG.M-follow-3PL  
 'They are happy.'

In (49), *won* functions as the object of the verb in an idiom meaning 'take a deep breath' (literally 'pulls heart hard'); in this idiom, the verb obligatorily occurs with masculine object inflection, agreeing with *won*.



- (49) *Kum won m-ekēle-n tetiet.*  
 1SG heart 1SG-pull-3SG.M hard  
 ‘I took a deep breath.’

Another word that is feminine in its literal meaning but masculine in idioms is *puna* ‘brain’. In (50), *puna* controls feminine subject agreement in its literal meaning, while in (51), it controls masculine object agreement in an idiom *-ekelen puna* ‘to snore’ (literally ‘to pull one’s brain’).<sup>10</sup>

- (50) *Kum puna w-o cheliel.*  
 1SG brain 3SG.F-be hot  
 ‘My brain hurts.’
- (51) *Chi n-ekēle-n puna kisiel.*  
 2SG 2SG-pull-3SG.M brain fast/loud  
 ‘You were snoring loudly.’

A final instance of a word that is obligatorily masculine is the interrogative pronoun *mon* ‘who’, illustrated in (52). It is not possible to use a verb form *chaltawro* in (52), with 3SG.F object agreement, even in contexts where it is assumed that someone is looking for a woman, although 3PL agreement would be possible if it is assumed that more than one person is being looked for.

- (52) *Chim ch-altawro-n mon?*  
 2PL 2PL-look-3SG.M who  
 ‘Who are you looking for?’

*Mon* thus behaves as a masculine noun.<sup>11</sup>

<sup>10</sup>Note that in all the examples I have discussed where a noun is a different gender in an idiom from its gender outside of idioms are cases where the noun is masculine in the idiom but feminine outside of idioms. This appears to be due to the fact that the relevant nouns denote inanimate objects outside of idioms and thus are feminine outside of idioms.

<sup>11</sup>There is no interrogative pronoun in Walman meaning ‘what’. Rather, there is an interrogative adnominal word *mol* and the expression for ‘what’ is *opucha mol* literally ‘what thing’. The gender of noun phrases with *mol* is determined by the gender of the noun (or the sex of the referent).

## 4 Pluralia tantum nouns

I analyse nouns in Walman which are always grammatically plural as pluralia tantum nouns (Corbett 2012: 233ff; Acquaviva 2008).<sup>12</sup> While the category of pluralia tantum nouns in other languages is not usually considered a gender, what makes it gender-like in Walman is the sheer number of pluralia tantum nouns. In our current data, there are about twice as many pluralia tantum nouns as there are masculine nouns.<sup>13</sup> What this means is that apart from nouns which can be either masculine or feminine depending on the sex of the referent, every noun in Walman is masculine, feminine, or pluralia tantum. In this sense, pluralia tantum is like a gender.

In many languages, what characterizes pluralia tantum nouns is that they are plural in form (e.g., *scissors* in English). In Walman, however, what characterizes pluralia tantum nouns is not their form, but the fact that they always trigger plural agreement. An example of a pluralia tantum noun is *nyi* ‘fire’. In (53), it triggers 3PL subject agreement on the verb *yiri* ‘stand up, rise’ and *yreliel* ‘shine, for a fire to blaze’.

- (53) *Nyi y-iri pa, nyi y-reliel.*  
 fire 3PL-stand.up PTCL fire 3PL-shine  
 ‘The fire rose, it was ablaze.’

In (54), the same noun triggers 3PL object agreement on *noysusur* ‘move’ and 3PL subject agreement on *yesei* ‘go outside’.

- (54) *Runon n-o<y>susur nyi y-esi chalien.*  
 3SG.M 3SG.M-move<3PL> fire 3PL-go.outside outside  
 ‘He moved the fire outside.’

And in (55), the same noun triggers 3PL object agreement on the verb *kaoy* ‘shoot’ (here used in the sense of ‘light’ in ‘light a fire’), as well as plural agreement on the numeral *ngony* ‘one’.

<sup>12</sup>Many linguists distinguish a singular expression *plurale tantum* from a plural expression *pluralia tantum*. But there is considerable inconsistency in the literature in the use of these expressions, so I avoid the expression *plurale tantum* and urge other linguists to do likewise. In this paper, I treat the expression *pluralia tantum* as grammatically similar to the words *masculine* and *feminine*.

<sup>13</sup>Our current data includes 81 instances of pluralia tantum nouns, but only 40 instances of masculine nouns. Since there are a number of nouns denoting animals whose gender we have not yet had opportunity to check, it is likely that the ratio of pluralia tantum nouns to masculine nouns will be less than 2 to 1.

- (55) *Kipin k-ao-y nyi ngo-ny.*  
 1PL 1PL-shoot-3PL fire one-PL  
 ‘We lit a fire.’

In (56), the pluralia tantum noun *apar* ‘platform, shelf, bed’ triggers plural agreement on the demonstrative *payten* and 3PL subject agreement on the verb *yo* ‘be’.

- (56) *Apar pa<y>ten y-o rachi.*  
 bed that<PL> 3PL-be strong  
 ‘That bed is strong.’

Just as there are semantic factors that partially account for gender in Walman, there are also semantic factors that probably account for at least some pluralia tantum nouns in Walman. Like pluralia tantum nouns in many languages, there is something about many pluralia tantum nouns in Walman that can be conceived as denoting more than one thing. In the case of *nyi* ‘fire’, there are multiple flames. In the case of *apar* ‘bed, shelf’, there are multiple pieces of wood. Other pluralia tantum nouns that denote objects that contain multiple pieces of wood include *chauchau* ‘door’, *salriet* ‘steps’, and *watakol* ‘raft, coffin’. Pluralia tantum nouns that contain multiple threads (or similar material) include *chrikiel* ‘net’, *ranguang* ‘clothes’ and *kmem* ‘rope for tying logs together to form a raft’. The noun *tim* ‘dew’ is pluralia tantum and could be construed as involving multiple drops. The noun *yikiel* ‘language, story, statement, word’ is pluralia tantum and one could think of most of these uses as involving multiple words.

However, there are many nouns that can just as easily be conceived of as denoting something with multiple pieces that are not pluralia tantum nouns, including *yie* ‘bilum, string bag’, *wuwu* ‘basket made from spines of nipa palm fronds for trapping fish’, and *amen* ‘type of basket made from coconut leaves, used for fishing’. Conversely, there are pluralia tantum nouns where it is less obvious that they consist of multiple instances of something, such as *nganyi* ‘urine’, *almat* ‘fog’, *ei* ‘lime (white powder produced from grinding up shells, used when chewing betelnut)’. All three of these nouns are mass nouns, but mass nouns do not appear to be pluralia tantum nouns with any greater frequency than count nouns. For example, *wul* ‘water’ and *tantan* ‘sand’ are mass nouns, but are grammatically feminine (as illustrated for *wul* ‘water’ in (16) above by the feminine object agreement on the verb *nako* ‘eat’ and the feminine form of the demonstrative *paten*).

One of the more interesting classes of pluralia tantum nouns are ones denoting body parts. The majority of these nouns denote body parts that occur in pairs.

However, these nouns trigger plural morphology even when only one of the two parts is denoted, as in (57), where *chkuel* ‘eye’ triggers plural agreement on both *ngony* ‘one’ and *yo* ‘be’.

- (57) *Chi chkuel ngo-ny tu y-o ngul.*  
 2SG eye one-PL PERF 3PL-be blind  
 ‘One of your eyes is blind.’

Other pluralia tantum nouns denoting body parts that occur in pairs include *kam* ‘lungs’, *kayal* ‘foot’, *kawa* ‘heel’, *kopun* ‘buttock’, *nyiminy* ‘breast’, *wi* ‘palm of hand, hand not including fingers’, *mkuel* ‘ear’, and *wili* ‘shoulder’. However, some pluralia tantum nouns refer to body parts that are not normally regarded as paired, such as *repicha* ‘mouth’, *chpurum* ‘upper lip’, *saykil* ‘liver’, *ngoul* ‘womb’ and *kal* ‘afterbirth’. There are also some body part nouns in Walman which occur in pairs but which are not pluralia tantum nouns; however, in each case, these are nouns that have distinct plural forms, such as *kampotu* ‘knee’ (plural *kamtikiel*).

Note that while pluralia tantum nouns can be conceived of as denoting things with multiple parts, they can still denote single objects, that is, single objects with multiple parts. In other words, they can be semantically singular, as reflected by the fact that they can be modified by either of two words meaning ‘one’ with plural inflection, as in (58) and (59), as well as (31), (55) and (57) above.

- (58) *Kum ranguang alpa-ny.*  
 1SG clothing one-PL  
 ‘I have one shirt.’
- (59) *Kum m-oko-y chrikiel ngo-ny.*  
 1SG 1SG-take-3PL net one-PL  
 ‘I brought one net.’

Some nouns are optionally pluralia tantum. For example, the noun *tokun* ‘knot’ can be used with singular agreement to denote a single knot, but with plural agreement to denote either a single knot or more than one knot. Some nouns are pluralia tantum with one sense, but not with another. For example, the noun *wukul* denotes either the sail of a boat or the soft bark flap of a coconut tree, which is like a cloth and which is used to strain the sago dust out of the water in making sago. It is pluralia tantum with the first of these senses, but not with the second. A more complex example is illustrated by the noun *kiri*, which means either ‘sago flour’ or ‘sago pancake’. On the first of these meanings, it is optionally pluralia

tantum, while on the second it is always pluralia tantum. This is particularly interesting since it is semantically a mass noun with the first sense, but a count noun with the second; one might have expected it to be more likely pluralia tantum when a mass noun.

In the preceding section, I described a few nouns which are masculine in certain idioms but feminine outside of idioms. We are also aware of at least one case of a noun which does not occur outside of idioms, but which is feminine in one idiom but pluralia tantum in two other idioms. The word *apum* combines with *kakol* ‘skin’ to mean ‘body’, as in (60), where *loyol apum kakol wru* ‘a sugar-glider’s body’ triggers feminine agreement on the verb *wo* ‘be’.

- (60) *Loyol apum kakol w-ru w-o nngkal-nngkal, chei*  
 sugar.glider body skin GEN-3SG.F 3SG.F-be small-small tail  
*w-ru ro-ø rani.*  
 GEN-3SG.F piece-F long  
 ‘A sugar-glider’s body is small but its tail is long.’

However, the same word *apum* occurs in two idioms where it behaves as a pluralia tantum noun, controlling plural subject agreement on the verb. One of these idioms, *apum yo sopuer* ‘to feel tired’, is illustrated in (61), while the other, *apum yo mayay* ‘to feel ashamed’, is illustrated in (62).<sup>14</sup>

- (61) *Kum apum y-o sopuer.*  
 1SG body 3PL-be tired  
 ‘I am feeling lethargic.’
- (62) *Runon apum y-o mayay.*  
 3SG.M body 3PL-be shy  
 ‘He feels ashamed.’

<sup>14</sup>The adjectives *sopuer* ‘tired’ and *mayay* ‘ashamed’ can also be used with the experiencer as subject, as illustrated in (i) for *sopuer* ‘tired’.

- (i) *Kum m-o sopuer.*  
 1SG 1SG-be tired  
 ‘I’m tired.’

We do not know if there is a difference in meaning between these non-idiomatic uses of these adjectives and the idioms in (61) and (62).

The idiomatic uses in (61) and (62) involve psychological states while the use in (60) does not. This is probably not a coincidence since the idioms in (61) and (62) resemble the idioms in (47) and (48), where the noun *won* ‘heart’ controls masculine agreement and the meaning involves psychological states.

There are also a few nouns which are singularia tantum nouns that do not appear to be mass nouns. One such noun is *woru* ‘mosquito’, which always triggers feminine singular agreement, as in (63), where it controls feminine singular subject agreement on the verb *wanpu* ‘attack’.

- (63) *Kon woru chomchom w-a<n>pu.*  
 night mosquito many/much 3SG.F-attack<3SG.M>  
 ‘At night, many mosquitoes bit him.’

While examples like (63) are consistent with *woru* being a mass noun, the meaning of (64), where *woru* functions as object of *mkawlo* ‘count’, but still triggers singular agreement, implies that it is a count noun.

- (64) *Kum m-kawlo-ø woru.*  
 1SG 1SG-count-3SG.F mosquito  
 ‘I counted the mosquitoes.’

While pluralia tantum in Walman behaves in some ways like a gender, I make no claim that it *is* a gender, though I am not aware of any strong arguments against this position. Note that if we were to consider pluralia tantum a gender, I would not be suggesting that plural is a gender, only that the forms used with pluralia tantum nouns are the same as those used for all plurals regardless of gender. A more detailed description of the kinds of nouns that are often pluralia tantum in Walman is given in Dryer (n.d.).

## 5 Diminutive

In this section, I describe the Walman diminutive, illustrated in (7) above, and discuss ways in which it is both like and not like a gender.<sup>15</sup> Corbett (2012: 149) argues that the Walman diminutive is indeed a gender, though a non-canonical one. In Dryer (2016), I discuss possible reasons not to consider it a gender.

<sup>15</sup>My discussion in this section is brief since I discuss the Walman diminutive in more detail in Dryer (under revision) and Dryer (2016).

Unlike diminutives in most languages, the Walman diminutive is inflectional (rather than derivational) in that diminutive affixes occur in the same morphological positions as affixes coding gender and number. In (65), for example, we get diminutive subject prefixes on the verbs *lan* ‘be at’ (here functioning as a progressive auxiliary verb) and *loruen* ‘cry’.

- (65) *Nyanam nngkal pa l-an l-oruen.*  
 child small that 3.DIMIN-be.at 3.DIMIN-cry  
 ‘The small child was crying.’

And in (66), we get diminutive agreement on the demonstrative *palten*, on the verb *lo* ‘be’ and on the adjective *lapol* ‘large’.

- (66) *Pelen pa<I>ten l-o lapo-l.*  
 dog that<DIMIN> 3.DIMIN-be large-DIMIN  
 ‘That puppy is large.’

All words that can inflect for gender and number can also inflect for diminutiveness.

What makes diminutive significantly different from masculine and feminine gender is that there are no nouns that are lexically diminutive, that is, there are no nouns which obligatorily trigger diminutive agreement.<sup>16</sup> In principle, any noun can be associated with diminutive agreement. For example, the noun *chu* ‘wife’ is normally feminine, but in (67), it triggers diminutive subject agreement on the verb *lalma* ‘die’ in the relative clause *ni lalma pa* ‘who died there’ modifying *chu*.

- (67) *Runon n-akrowon chu ni l-alma pa.*  
 3SG.M 3SG.M-think wife REL 3.DIMIN-die there  
 ‘He mourned his dear wife who had died there.’

The semantics associated with the Walman diminutive is similar to the semantics associated with derivational diminutives in other languages. It can simply denote a smaller size than normal, as in (68), where it triggers diminutive object agreement on the verb *malwul* ‘buy’.

<sup>16</sup>There is one word that may be (or may be considered) lexically diminutive that I discuss in Dryer (under revision), viz. *kamtel*, the diminutive form of *kamten* ‘man’. However, as discussed in Dryer (under revision), there are reasons to consider this the diminutive form of a single lexical item rather than a distinct lexical item.

- (68) *Kum m-a<I>wul                    selenyue.*  
       1SG    1SG-buy<3.DIMIN> axe  
       ‘I bought a small axe.’

However, it more often denotes the young of a species, as in (65) and (66) above, or expresses endearment, as in (67) above.

Apart from the fact that there are apparently no lexically diminutive nouns in Walman, another reason for thinking that the Walman diminutive is not a gender is that one can get agreement mismatches in the sense that one target of agreement for a given controller is masculine or feminine while another target of the same controller is diminutive, suggesting that a given noun phrase can be masculine or feminine but at the same time diminutive. For example, in (69), the noun phrase *wuel woyuel* ‘the naughty pig’ is masculine, triggering masculine subject agreement on the verb *narul* ‘run away’, but at the same time diminutive in that the adjective *woyuel* ‘bad’ exhibits diminutive inflection.

- (69) *Wuel woyue-I    n-arul.*  
       pig    bad-DIMIN 3SG.M-run.away  
       ‘The naughty little male pig ran away.’

The reverse is also possible, with masculine inflection on the adjective and diminutive agreement on the verb, as in (70).

- (70) *Wuel woyue-n    I-arul.*  
       pig    bad-MASC 3.DIMIN-run.away  
       ‘The naughty little male pig ran away.’

Whether the Walman diminutive should be treated as a gender is a complex question and depends to a large extent on how one interprets the question, as discussed by Dryer (2016). For more detailed description of the Walman diminutive, see Dryer (under revision) and Dryer (n.d.).

## 6 Conclusion

In this paper, I have described gender in Walman. The choice between the two clear instances of gender, masculine and feminine, is largely predictable semantically, though this is partly due to the fact that inanimate nouns are always feminine. The only nouns whose gender is apparently arbitrary are ones denoting animals. I have also briefly described two other gender-like phenomena in



Walman, pluralia tantum and diminutive. I do not take a stand here on whether these two phenomena are genders or not. My goal has simply been to illustrate ways in which they are gender-like and ways in which they are not gender-like. In the case of pluralia tantum nouns, they are more gender-like than similar categories in other languages, simply because there are so many of them. In the case of the diminutive, it is like a gender to the extent that it is coded in the same morphological positions as masculine and feminine, but not like a gender in that there appear to be no lexically diminutive nouns.

## Acknowledgments

I acknowledge funding supporting field work by myself and Lea Brown on Walman from the Endangered Languages Documentation Programme and from the National Science Foundation (in the United States). We began field work on Walman in 2001 and are currently preparing a detailed description of the language (Dryer n.d.). See Brown & Dryer (2008) for some basic features of Walman. I am indebted to anonymous reviewers and particularly to Lea Brown for comments on an earlier draft of this paper.

## Special abbreviations

The following abbreviations are not found in the Leipzig Glossing Rules:

DIMIN	diminutive	PTCL	particle
RECIP	reciprocal	PERF	perfect
REL	relative clause marker	Q	marker of polar question

## References

- Acquaviva, Paolo. 2008. *Lexical plurals: A morphosyntactic approach*. Oxford: Oxford University Press.
- Brown, Lea & Matthew S. Dryer. 2008. The verbs for ‘and’ in Walman, a Torricelli language of Papua New Guinea. *Language* 84(3). 528–565.
- Corbett, Greville G. 2012. *Features*. Cambridge: Cambridge University Press.
- Dixon, Robert M. W. 1977. Where have all the adjectives gone? *Studies in Language* 1. 19–80. [Reprinted 1982 in R. M. W. Dixon: *Where have all the adjectives gone and other essays in syntax and semantics*, pp. 1–62. Berlin: Mouton.]
- Dryer, Matthew S. N.d. *A grammar of Walman*.

Dryer, Matthew S. 2016. Crosslinguistic categories, comparative concepts, and the Walman diminutive. *Linguistic Typology* 20(2). 305–331.

Dryer, Matthew S. under revision. Diminutive as an inflectional category in Walman. Currently being revised for resubmission to *Oceanic Linguistics*.

## Chapter 8

# The gender system of Coastal Marind

Bruno Olsson

Australian National University

The gender system of Coastal Marind (a Papuan language of the Anim family of South New Guinea; Usher & Suter 2015) is treated in relative detail in Drabbe's (1955) masterful grammar. The division of nouns into four genders (basically masculine, feminine and two inanimate genders) is familiar from various languages around the globe, but the morphology of exponence (gender agreement marked to a large extent by stem-internal changes on targets) is somewhat more exotic and is occasionally cited in the literature. In this paper I provide an overview of the system, combined with discussion of two issues: the origins of stem-internal gender agreement, and the wide-ranging syncretism between animate plurals and the 4th gender (the 2nd inanimate gender). I show that this 'syncretism' makes the status of the 4th gender ambiguous, since the members of this gender also could be analysed as an unusually large class of pluralia tantum. While I argue that the synchronic 4-gender analysis must be maintained for Coastal Marind, I speculate that an erstwhile grouping of pluralia tantum provided the diachronic source of the 4th gender.

**Keywords:** Gender, number, morphology, diachrony, Papuan languages.

## 1 Introduction

The idea that gender systems can become more complex (add a gender or two) through the 'reinterpretation' of some non-gender feature as signalling a gender value has a long history in linguistics (e.g. Brugmann 1891 on the origins of the Indo-European feminine gender). In this paper I show that the fourth gender of Coastal Marind could be more parsimoniously described as pluralia tantum in a 3-gender system; however, I will argue that semantic considerations ultimately force us to retain the traditional four-gender description.



Based on its ambiguous status in Coastal Marind, I will speculate that the fourth gender in the languages of the Anim family of South New Guinea could have originated as a grouping of pluralia tantum nouns, and that subsequent changes in the agreement system and attraction of additional nouns to the emerging fourth gender could have led to a present situation where the pluralia tantum analysis is no longer possible, resulting in a 4-gender system.

I also add further support to Usher & Suter's (2015) proposal that one of the main manifestations of gender agreement in the language – stem internal vowel alternations in agreement targets – arose from a process of umlaut triggered by postposed articles, by showing that the synchronic distribution of stem-final vowels in nouns is consistent with gender umlaut affecting a much larger part of the lexicon than just present-day gender-agreeing lexemes. The discussion is based on data from the best known Anim language, Coastal Marind (for a modern reference grammar, see Olsson 2017).

The article is structured as follows. §1.1 is a brief demonstration of the four genders of Coastal Marind. The language is placed in its areal and genealogical context in §1.2, while §1.3 provides information about some relevant structural features of Coastal Marind. §2 describes the interesting correlation between stem-final vowels and gender membership in nouns, showing that it is of limited productivity synchronically, but likely derives from an earlier system of post-nominal gender articles. §3 describes gender agreement across the clause, with emphasis on the systematic correspondence between exponents of Gender IV and the plural of Gender I/II. §4 shows that this correspondence continues in the participant indexing on the verb. This suggests an alternative analysis according to which Gender IV is an unusually large group of pluralia tantum rather than a gender of its own. In §5 I will show that the assignment of nouns to Gender III and IV is largely arbitrary, but that the occurrence in Gender IV of many nouns that are typical pluralia tantum nouns across languages is suggestive of being a remnant of such a grouping. I also show that a similar pattern occurs in Mian, a language that probably is a distant relative of Coastal Marind since the Anim and Ok families (to which Mian belongs) are likely members of the enormous Trans-New Guinean super-family. I conclude that the 4-gender analysis should be maintained for the present state of Coastal Marind, but that the pluralia tantum nouns possibly provided the source for the fourth gender.

### 1.1 The Coastal Marind 4-gender system

The existence of a 4-gender system in Coastal Marind is evident if one compares the form of the demonstrative *Vpe* (where *V* stands for a vowel) or the adjec-

tive *samlayVn* ‘mid-size, neither big nor small’ combined with different nouns in examples (1)–(3). As indicated by the hyphens, attributively used adjectives are compounded with their head nouns. The nouns themselves are invariant.

- (1) a. *samlayen-patul e-pe*  
mid.size:I-boy(I) I-that
- b. *samlayun-kyasom u-pe*  
mid.size:II-girl(II) II-that  
‘that mid-size boy/girl’
- (2) a. *samlayin-patul i-pe*  
mid.size:I/II.PL-boys(I) I/II.PL-that
- b. *samlayin-kyasom i-pe*  
mid.size:I/II.PL-girls(II) I/II.PL-that  
‘those mid-size boys/girls’
- (3) a. *samlayan-da e-pe*  
mid.size:III-sago(III) III-that  
‘that mid-size sago palm/those mid-size sago palms’
- b. *samlayin-bomi i-pe*  
mid.size:IV-termite.mound(IV) IV-that  
‘that mid-size termite mound/those mid-size termite mounds’

All nouns denoting male humans behave like *patul* ‘boy’ (in 1a) in combining with a demonstrative with the initial vowel *e-* in the singular; nouns denoting female humans (and all animals) pattern like *kyasom* ‘girl’ (1b) in combining with an *u-* initial demonstrative. As the examples in (2) show, these nouns exhibit a contrast in number. The demonstrative has to be *ipe* in the plural, and the adjective, which is compounded with its head noun, has the exponent vowel *i* in the final syllable of the stem.

The nouns in (3) are inanimate, and trigger different vowels on the demonstrative: *da* ‘sago palm’ triggers *e-*, *bomi* ‘termite mound’ triggers *i-*. Note that the resulting forms are homophonous with demonstratives in the preceding examples: *epe* in (3a) with the demonstrative used for *patul* in (1a), and *ipe* in (3b) with the plural forms in (2). For (3a), the distinct form *samlayan* of the adjective proves that this is indeed a separate gender, although the agreement of the demonstrative happens to be homophonous with that seen in (1a). But the case

in (3b) is more difficult, since the agreement on both the demonstrative and the adjective turns out to be homophonous with the plural forms. I will return to this pervasive syncretism further below.

The four agreement classes – from now on referred to as Gender I, II, III and IV – are summarized in Table 1, as evidenced by the exponence pattern of *samlayVn*.

Table 1: Exponents of agreement on *samlayVn* ‘mid-size’

	SG	PL
I	e	i
II	u	
III		a
IV		i

These data represent one of the most well-known gender systems in New Guinea. The Coastal Marind system of four grammatical genders has featured in prominent publications such as Corbett (1991: 116) and Aikhenvald (2000: 60) after having been brought to the fore in Foley’s influential compendium on Papuan languages (Foley 1986: 82–83). This attention is due to the description of the gender system provided in Petrus Drabbe’s extensive grammar of the language (Drabbe 1955). Few researchers seem to have had the courage to dive deeper into Father Drabbe’s sometimes quite demanding *Spraakkunst*, so one purpose of this article will be to give a more representative picture of the gender system and its manifestations, and, in particular, the syncretism between animate plurals and Gender IV. The data come from my own fieldwork on the Western variety of Coastal Marind, a dialect that is mutually intelligible with the Eastern variety described by Drabbe.

## 1.2 Coastal Marind in context

The varieties collectively known as Coastal Marind are spoken in ca. 40 villages along the coast of the Arafura sea and in the adjoining swampy lowlands. I estimate the total number of speakers to be around 14.000 based on government and SIL figures. The Coastal Marind land forms part of the linguistically diverse Trans-Fly area (Evans 2012; Evans et al. 2018) straddling the border of present-day Indonesia (where Coastal Marind is spoken) and the independent country of Papua New Guinea.

The dialect situation is complex, and it is probable that ongoing research will show that some of the varieties described in the literature as dialects are in fact distinct languages. Dialectal variation in gender would likely be an interesting area to explore, as there are differences (mainly in assignment) even between villages speaking virtually identical varieties of Coastal Marind. On the whole, however, the basics of gender and agreement are the same in all known varieties, so the data presented here (from the village of Wambi) are representative of all coastal varieties, and probably of the (less well-known) inland varieties as well.

On a higher level, gender has recently emerged as a crucial factor in the genealogical classification of Coastal Marind. Usher & Suter (2015) show that gender ablaut in nouns such as *anem* ‘man’, *anum* ‘woman’ and *anim* ‘people’ recur throughout a number of languages of the Trans-Fly region. This observation, in addition to a large set of lexical cognates showing regular sound correspondences, leads Usher & Suter to propose a hitherto unrecognized language family – the Anim family, named after the recurring word for ‘people’ – of which Coastal Marind so far is the only language for which substantial descriptive work is available. Obviously, more work on the other Anim languages – several of which are rapidly losing speakers – could provide crucial insights into the development of the Anim gender system.

### 1.3 Typological background

Some of the structural features of Coastal Marind are relevant to the description of its gender system. Coastal Marind displays the relatively rare combination of verb-final constituent order and massively prefixing verb inflection. Based on co-occurrence, a prefixal template with ca. 18 slots can be set up, marking notions such as tense, various aspectual distinctions, applicatives, reciprocal, various adverbial meanings (‘again’, ‘first’, ‘far away’, ‘in contact with surface’) and indexation of (roughly) actor, recipient and affected possessor; undergoer indexation is in turn marked on the verb stem by complicated alternations including pre-, suf-, in-, and circumfixal morphology.

Some of the prefixes occupying the first (i.e. leftmost) positions agree in gender with an argument, although they primarily mark grammatical distinctions other than gender (e.g. tense-aspect). The prefixes devoted to argument indexing, on the other hand, reflect person and number but are insensitive to gender (with some exceptions to be discussed later). The verb stem itself is an important site for the manifestation of gender, so the intricate stem changes will be crucial to the arguments made here.

A relatively straightforward example of how verbs are segmented is given in (4). This verb has two prefixes, of which the first (leftmost) prefix agrees in gender with the subject (plural of Gender I/II). The stem is separated from the prefixal complex by a phonological boundary (indicated in glossing by means of a trailing hyphen followed by a blank). The formative *n-* on the stem marks it as the 1st person undergoer form, which clearly is a mismatch since there is no 1st person participant involved in the event. This idiosyncrasy is part of the reciprocal construction, and such value mismatches are not uncommon in Coastal Marind (cf. §4).

- (4) *ip-enam-                      n-asak-e*  
 ABSC:I/II.PL-RECP- 1.U-fight-IPFV  
 ‘They are fighting.’

Nominal morphology is sparse: there is no case marking and most nouns do not show overt gender marking. The exception is a handful of nouns (mostly kinship terms) that show alternations in the stem-final vowel according to gender (see below). This marking pattern also occurs on a subset of adjectives which agree with a noun in attributive and predicative use. The majority of adjectives are invariant and fail to show agreement. Instead, the main loci of gender agreement outside verbs are demonstratives and pronominal-like words (emphatic pronouns, question words). In the next section I turn to the reflexes of gender in nouns and what they can tell us about the diachronic development of gender marking in this part of the lexicon.

## 2 The manifestation of gender in nouns

### 2.1 Overt gender

A comparison of gender agreement across different word classes confirms that the picture emerging from examples (1)–(3) above is correct. All words that show morphological alternations according to gender follow these four agreement classes, although exponents vary across the targets showing agreement, and although many targets do not distinguish all four classes. Before dealing with agreement proper, we will consider nouns displaying OVERT GENDER. Whereas such alternations are not productive in contemporary Coastal Marind, a closer look reveals that traces of a more wide-ranging system of stem-final vowel alternations can be observed. The origins of this system of overt marking can be reconstructed following Usher & Suter (2015), as will be seen later.



Table 2: Overt gender on nouns

	I SG	II SG	I/II PL	III	IV
a.	<i>anem</i> 'man'	<i>anum</i> 'woman'	<i>anim</i> 'people'	<i>anem</i>	<i>anim</i>
b.	<i>namek</i> 'cousin (m)'	<i>namuk</i> 'cousin (f)'	<i>namik</i> 'cousins'		
c.		<i>namakud</i> 'animal'	<i>namakid</i> 'animals'	<i>namakad</i> 'thing(s)'	<i>namakid</i> 'thing(s)'
d.	<i>amnanggib</i> 'married man'		<i>amnangga</i> 'married men'		
e.	<i>wananggib</i> 'boy'	<i>wananggub</i> 'girl'	<i>wanangga</i> 'children'		
f.	<i>nahyam</i> 'my husband'	<i>nahyum</i> 'my wife'			
g.	<i>eyal</i> 'somebody (m)'	<i>eyul</i> 'somebody (f)'			
h.	<i>nanih</i> 'face (m)'	<i>nanuh</i> 'face (f)'	<i>nanih</i> 'faces'		

Some nouns with overt gender marking are listed in Table 2. Gender membership is reflected by the vowel in the final syllable of the stem (referred to as the 'stem-final vowel'), and the meaning of the noun is largely predictable from the gender. Thus, the skeletal stem *anVm* (a) can be thought of as having the general meaning 'person', which is narrowed down to 'man' when assigned to Gender I (*anem*), 'woman' in Gender II (*anum*), etc.; the stem *nahyVm* 'my spouse' (f) (*nahy* is a 1st person possessive prefix) giving 'husband' (*nahyam*, Gender I) and 'wife' (*nahyum* Gender II) once gender is assigned and vowels plugged into the stem.<sup>1</sup>

Assuming that the sets of gender forms derived from the skeletal stems are best treated as members of unitary lexemes, we can say that these lexemes are a proper subset of the nouns having REFERENTIAL GENDER (Dahl 2000), i.e. nouns that lack intrinsic gender and receive their gender value from the referent at

<sup>1</sup>Note that 'overt gender' only applies to nouns for which there is at least one other noun differing only in a stem-internal vowel, with a corresponding change in meaning. For example, the Gender IV noun *bomi* 'termite mound' does not have overt gender despite the presence of stem-final *i* (which is the general exponent of Gender IV agreement), since there are no corresponding nouns *\*bome*, *\*bomu* etc. to be found in the other genders.

hand. Most such nouns do not show overt gender, e.g. *yunayon* ‘infant’ (which takes agreement in Gender I or II depending on the sex of the referent).

The disassembly of Coastal Marind nouns into skeletal stems with inserted gender markers could appear to be a slightly misleading way of approaching the gender system of the language, since the phenomenon is fairly marginal. Only a dozen lexical items or so display the vowel alternation,<sup>2</sup> and many of the expected forms are irregular (e.g. plural of *wananggVb* is *wanangga* ‘children’, there is no plural \**wananggib*) or simply non-existent (e.g. there is no plural of *eyVl* ‘somebody’). The vowel alternation seems to be complete only for the stems *anVm* and *namakVd*: in addition to the person-denoting triplet man/woman/people, the former provides the forms *anem* and *anim* for inanimate denotanda in Gender III and IV respectively, for example in some compounds denoting fruits (*ambun-anem*, a *Syzygium* species in Gender III), while *namakVd* apparently can be used for non-rational entities (animals, things) of all genders except the masculine I.<sup>3</sup>

Looking at more nouns from Gender I and II, it seems clear that the pattern of alternating vowels showing gender membership is exception rather than rule. Nouns in Gender I denoting male humans also include *patul* ‘boy’, *ad* ‘father’, *manday* ‘wife’s elder brother, younger sister’s husband’ and so on; these nouns do not participate in any alternation with corresponding plural or female-denoting nouns. Person-denoting nouns in Gender II that likewise show no trace of overt gender are *kyasom* ‘girl’, *nikna* ‘son’s wife’, *ne* ‘mother’s brother’s wife’ etc.

Although overt gender is found only in a very small portion of the nominal lexicon, it should be noted that some of these nouns are high-frequency items, such as the words corresponding to the stem *anVm*, whose combined score makes them more frequent than any other noun in my corpus. Outside the noun inventory, stem-final vowel alternation plays an important role in common agreement targets such as the emphatic pronoun *anVp* (‘-self’), adjectives such as *papVs* ‘small’ and the postposition *IVk* ‘from’. This means that overt gender on nouns, and stem-final vowel alternation in general, is a common feature of Coastal Marind discourse, and obviously not as marginal as it would seem from a dictionary count alone.

<sup>2</sup>There are a handful of other nouns with overt gender in addition to the ones shown in the table. All of these denote humans of different age-ranks or societal roles that are more or less obsolete today, so the corresponding terms are falling out of use.

<sup>3</sup>In fact it seems that the stem *namakVd* ‘animal/thing’ can be used in Gender I: speakers reported that *namaked* can be used to refer to a male, although apparently with pejorative overtones, although I have never observed this in spontaneous speech.

A central claim of the comparative work in Usher & Suter (2015) is that the vowel alternations according to gender occur in languages throughout the Anim family, and that its origins can be reconstructed. Consider the forms *aneme(a)* ‘man’, *anumu* ‘woman’, *animi* ‘people’ from the related language Ipiko, another member of the Anim family. Usher & Suter argue that the stem-final vowel in *anVm* and other alternating stems is a residue of an earlier system of postnominal articles marking the gender of the noun, and they reconstruct expressions such as *\*anem=e* ‘the man’, *\*anum=u* ‘the woman’, *\*anim=i* ‘the people’ (2015: 114). In an earlier stage the noun was invariant and it was the presence of the gender article that triggered umlaut in the stem-final syllable (the shape of the invariant stem is beyond what can be reconstructed from the available data).

Usher & Suter’s hypothesis is plausible, especially as it refers to a well-known process leading to stem-internal vowel alternations (cf. Germanic umlaut giving English *mouse* and *mice* triggered by an earlier plural ending *\*-iz*). It can be added that some alternations are likely the result of more recent derivations involving gender-marking morphology. For example, the word *wayuklu* ‘girl’ and its plural *wayuklik* ‘girls’ are probably related to the postposition ‘from’ which has the forms *luk* and *lik* in the feminine and plural respectively, and which seems to be the source of many deverbal nominals in Coastal Marind (see Geurtjens 1933: 335 for the etymology; cf. *dahahiplik* ‘drunkards’ from *dahahip* ‘become drunk (plural subject)’). However, the ultimate source of the vowel alternation in *IVk* ‘from’ is likely not distinct from the umlaut process giving rise to the forms of *anVm*, so the suggestion that some cases of synchronic vowel alternations are of more recent origin than the original umlaut is not intended as a counterexample to Usher & Suter, but as an indication that the alternating pattern propagated indirectly through the lexicon as a result of derivation.

## 2.2 Simulating the effects of umlaut in the lexicon

Given the observations of alternating nouns showing overt gender, and Usher & Suter’s suggestion that the alternation came about because of umlaut triggered by a postposed article, the following interesting question arises: are there traces of umlaut also in non-alternating noun stems?

If umlaut was a regular process, we would expect it to have appeared with many nouns, as long as they were used with postposed articles. In the ideal case, all nouns in Gender I would have ended up with the stem-final vowel *e*, those in Gender II stem-final *u*, Gender III *a*, and those in Gender IV *i*. This is clearly not the case, as shown by the counts of stem-final vowels in Table 3. The table displays the frequency with which each of the five vowels of Coastal Marind occurs

in the last syllable of nouns whose gender membership has been determined. I have excluded all nouns showing overt gender from the counts, since we already know that their stem-final vowels correlate with gender membership. This is the reason why Gender I has so few members: the remaining male-denoting nouns have overt gender (e.g. *anVm*). Gender II likewise contains only a handful of female-denoting nouns, but has a higher count since it includes all names of animals.

Table 3: Distribution of stem-final vowels in nouns according to gender

	I ( <i>e</i> )	II ( <i>u</i> )	III ( <i>a</i> )	IV ( <i>i</i> )	Tot.
/i/	5	29	25	44	103
/u/	0	27	39	19	85
/e/	1	15	31	13	60
/o/	2	22	34	14	72
/a/	4	55	108	29	196
Tot.	12	148	237	119	516

Consider now the possibility that stem-final vowels of nouns and gender membership correlate to some degree, despite there being no one-to-one match. We are particularly interested in the vowels *e*, *u*, *a* and *i*, which Usher & Suter (2015) identify as the vowels of the proto-Anim demonstrative.<sup>4</sup> The vowels are given inside parentheses after their associated genders at the top of the table. We cannot test the correlation for Gender I, since there are too few nouns assigned to this category. The relevant cells for the remaining three genders have been shaded in Table 3. We now need to ascertain whether these scores could have been produced by a chance distribution of stem-final vowels, or whether they are non-random, thereby providing evidence that the umlaut pattern is found beyond the synchronically attested overt gender nouns.

To test this, I performed a simulation in which the nouns were reassigned randomly to the four genders (keeping the proportions intact), and then counted the frequency with which the vowels turned up in each gender. This procedure was then repeated a total of 200.000 times; the accumulated counts for the occurrence of the relevant vowels in Gender II, III and IV are presented in Figure 1, with the actual frequency of the vowel represented by the cross on the x-axis. The results

<sup>4</sup>In fact, Usher & Suter (2015: 119) tentatively reconstruct both *\*a* and *\*o* for the proto-Anim Gender III, but the exponent *o* is rare in Coastal Marind.

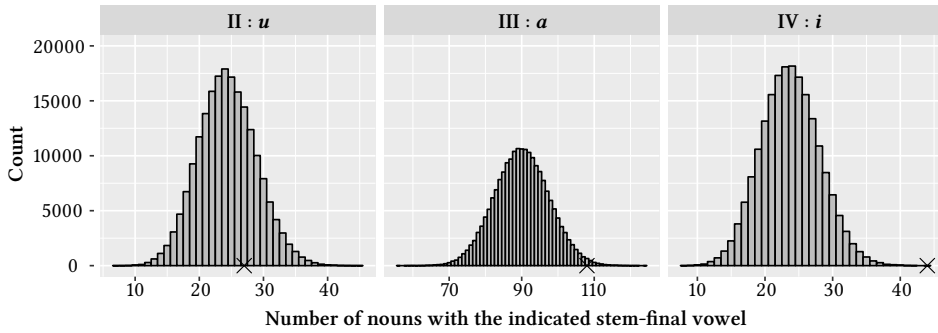


Figure 1: Actual and simulated distributions of stem-final vowels

show that two of the vowels are over-represented to a significant degree: *a* as the stem-final vowel in Gender III ( $z=2.40$ , adjusted  $p<0.05$ ) and *i* as the stem-final vowel of Gender IV ( $z=4.65$ , adjusted  $p<0.001$ ). These results support the hypothesis that gender umlaut affected a part of the lexicon that is larger than the set of nouns with overt gender, including many nouns of Gender III and IV.

No other positive skewings were close to statistical significance. This is somewhat surprising for Gender II, which would be expected to show a preference for *u* as the stem-final vowel (cf. the leftmost pane in Figure 1). I have no explanation for this, but it is worth noting that Coastal Marind seems to differ from other Anim languages in the uniform assignment of animals to Gender II: animals turn out to be divided between Gender I and II (the ‘masculine’ and ‘feminine’ genders) in Kuni (Edwards-Fumey 2007: 9), Ipiko (Usher & Suter 2015: 117, examples 16–17), and Bitur (Phillip Rogers, pers. comm.) which belong to three distinct sub-branches of Anim. A possible scenario would be that the reassignment of all animals to Gender II is an innovation present in Coastal Marind, which then would have obliterated any preponderance of *u* in Gender II as the new members entered.

### 3 Gender agreement

I will now consider how gender is manifested across agreeing pronominals, demonstratives and adjectives.<sup>5</sup> The purposes will be to give an overview of the agreement system, which contains some typologically interesting features, and more specifically to show that the apparent syncretism noted above between

<sup>5</sup>There is one more type of agreement target, viz. the four postpositions *IVk* ‘from’, *nV* ‘without’, *tV* ‘with’ and *hV* ‘like’. They are interesting for a variety of reasons, but I omit them from discussion here.

Table 4: Pronominal and demonstrative targets

Gloss	I SG	II SG	I/II PL	III	IV
‘whats-his/her-name, whatchamacallit’	<i>age</i>	<i>agu</i>	<i>agi</i>	<i>ago</i>	<i>agi</i>
‘who/what’	<i>ta</i>	<i>tu</i>	<i>ti</i>	<i>ta</i>	<i>ti</i>
‘him-/her-/itself/themselves’	<i>anep</i>	<i>anup</i>	<i>anip</i>	<i>anep</i>	<i>anip</i>
‘this/these’	<i>ehe</i>	<i>uhe</i>	<i>ihe</i>	<i>ehe</i>	<i>ihe</i>

Gender IV and the plural of Gender I/II is observed throughout the system. It even turns up in some unexpected places, prompting the question of whether the system is not better analyzed as comprising three genders instead of four, a possibility that will be further explored in §4, §5 and §6.

### 3.1 Pronominals and demonstratives

The only word classes in which agreement is found on a majority of the members are demonstratives and pronominals. Agreement on the distal demonstrative *Vpe* was seen in (1)–(3) above; some more examples of agreeing targets within these categories are in Table 4. While the small set of personal pronouns in Coastal Marind (*nok* ‘I, we’ *oy* ‘2SG’, *yoy* ‘2PL’) show no gender distinction, gender agreement is pervasive across other pronominal-like elements such as question words (e.g. *tV* ‘who, what’ *Vn* ‘where, which’) and the polyfunctional word *agV*, which has among its uses that of a placeholder ‘whats-his/her-name’ (referring to a person) or ‘whatchamacallit’ (referring to a thing).<sup>6</sup> Note that, in contrast to the various unpredictable exponents of Gender I and III, the exponents of Gender II (*u*) and Gender IV (*i*) are constant across all targets, with the latter showing homophony with the I/II plural in all four items.

### 3.2 Adjectives

Coastal Marind adjectives are similar to nouns in that both classes lack the luxuriant inflectional possibilities of verbs. The main morphosyntactic feature distin-

<sup>6</sup>Forcing speakers to choose a gender for words meaning ‘who, what?’ that refer to some unknown entity might seem counter-intuitive since the gender of the referent must be unknown in many cases (since there is no clear semantic basis for Gender III and IV); cf. European languages restricting gender agreement to attributive ‘which’ (e.g. Russian *kotoryj* ‘which (masc.)’ etc.) while pronominal ‘who’ lacks agreement (e.g. Russian *kto* ‘who’). Gender agreement on placeholders appears more common, especially in placeholders of phrasal and/or pronominal origin such as English *whatchamacallit* etc.

Table 5: Gender agreement on adjectives

Gloss	I SG	II SG	I/II PL	III	IV
‘light (weight)’	<i>akek</i>	<i>akuk</i>	<i>akik</i>	<i>akak</i>	<i>akik</i>
‘short’	<i>dahwages</i>	<i>dahwagus</i>	<i>dahwagis</i>	<i>dahwagis</i>	<i>dahwagis</i>
‘thin’	<i>halahel</i>	<i>halahul</i>	<i>halahil</i>	<i>halahal</i>	<i>halahil</i>
‘sharp’	–	–	–	<i>yayayaŋ</i>	<i>yayayiy</i>
‘dull’	–	–	–	<i>yandayal</i>	<i>yandayil</i>
‘old, ancient’	<i>taname</i>	<i>tanamu</i>	<i>tanami</i>	<i>tanama</i>	<i>tanami</i>
‘strong’	<i>tage</i>	<i>tagu</i>	<i>tagi</i>	<i>taga</i>	<i>tagi</i>
‘ripe’	–	–	–	<i>eho</i>	<i>ihu</i>

guishing adjectives from nouns seems to be the lack of inherent gender. A small subclass of adjectives (13 members are known in the Western dialect) agree in gender, some of which are shown in Table 5. Other adjectives are invariant (e.g. *yaba* ‘big’, *ndom* ‘bad’, *waninggap* ‘good’). The patterns of exponence largely follow those familiar from nouns with overt gender, with agreement marked by means of changes in the stem-final vowel, except for *VhV* ‘ripe’ which shows a unique pattern of vowel height harmony. Note that some of the adjectives are semantically incompatible with animates, whence the dashes in the table.

The forms of agreeing adjectives are much more regular than nouns with overt gender: Gender I and II consistently have /e/ and /u/ as their exponents, and their plural indicated by /i/; for inanimates, Gender III is largely indicated by /a/, while the pattern of homophony between the I/II plural forms and the Gender IV forms is observed again.

A remarkable exception from these regularities is the adjective ‘small’, whose forms are given in Table 6. This adjective is noteworthy for two reasons. First, it is the only word in the language that distinguishes singular and plural for Gender III and IV. This is done by means of the suppletive stems *isahih* and *wasasuy*, neither of which bear any phonological resemblance to the singular stem *papVs*. Following Corbett (1991: 168) we can say that ‘small’ is OVER-DIFFERENTIATED since it distinguishes a feature (number of inanimates) which is absent elsewhere in the system. However, one could also argue that ‘small’ does not show true agreement for gender, because the stems involved are suppletive. This is the approach taken by Durie (1986: 362), who – speaking of verbal number suppletion – argues that “suppletive stems select for rather than agree with the number of their argument”. Either way we look at it, ‘small’ has to be marked as an exceptional item,

and does not detract from the generalization that number as a nominal category is restricted to the animates, e.g. the members of Gender I and II.

Table 6: Gender agreement on ‘small’

	I	II	III	IV
SG	<i>papes</i>	<i>papus</i>	<i>papes</i>	<i>papis</i>
PL	<i>isahih</i>	<i>isahih</i>	<i>wasasuy</i>	<i>isahih</i>

Second, the stems used for ‘small’ in the plural are *isahih* and *wasasuy*, of which the former (which is also used as a noun meaning ‘children, young of animals’) is used not only for animates, but also for plural of Gender IV. This would be quite surprising if the syncretism between I/II plural and Gender IV noted so far (e.g. the demonstrative *ipe* covering I/II plural and IV) were merely a case of accidental homophony. Below we will see other cases where syncretisms between I/II plural and IV suggest a more profound relationship between the forms.

## 4 Agreement and participant indexing on verbs

The morphology of the Coastal Marind verb is complicated, and nominal gender plays a role within three of the inflectional sites of the verb: in a set of gender-agreeing prefixes, in the person indexing reflecting an UNDERGOER argument, and, somewhat marginally, in the indexing of the ACTOR argument of the verb. The gender-agreeing prefixes are the most straightforward, and behave largely like the non-bound agreeing items that we have seen so far. I will give some examples of gender agreement on the verb below. I contrast gender AGREEMENT with bound person marking on the verb, which I refer to as INDEXING. I will show below that these two phenomena behave quite differently in Coastal Marind, so it is convenient to make the terminological distinction between agreement and indexing in the description of the Marind verb.

Several inflectional prefixes are sensitive to the gender of some argument of the verb, although their main function lies in some other domain (e.g. tense-mode-aspect) so it is not appropriate to call them ‘gender prefixes’; rather, they are prefixes of which a sub-string happens to show agreement in gender. Let us take the prefix *Vp-* ‘ABSconditive’ as an illustration. Simplifying matters drastically, we can say that this prefix is used when the speaker is drawing attention to



some present state-of-affairs that is unavailable to the addressee, either because her attention is on something else, as in (5), or because she made a previous statement contradicting the state-of-affairs that actually holds, as in (6). The question of what argument of the verb controls the gender agreement in the prefixes is complicated, and I will not explore it here. Suffice to note that it is the (intransitive) subject in (5) that is the controller, whereas the Gender I agreement in (6) corresponds to the male recipient-like participant (other constellations would behave differently).

- (5) (Addressee standing facing away:)  
*kosi-awe up-Ø- kwayita!*  
 small-fish(II) ABSC:II-3SG.A- be.swimming.inside  
 ‘A little fish is swimming in there!’
- (6) (Reply to “You should talk to him!”, female speaker:)  
*ep-ak-o- lay-e!*  
 ABSC:I-1.A-3SG.DAT- talk-IPFV  
 ‘I am talking to him!’

Morphologically these prefixes are straightforward, since they have the same forms as the distal demonstrative *Vpe* (betraying a historical relationship), minus the final *-e*. The same holds, for example, for the continuative prefix *anVpand-* which most likely derives from the emphatic pronoun series *anVp* (cf. Table 4). Gender agreement in the prefixal complex then seems to be of relatively recent origin, resulting from the integration of free demonstrative and pronominal elements into the verb. Once more, the syncretism between the Gender I/II plural and Gender IV that was encountered in the nominal targets recurs in the prefixal agreement, so the Absconditive prefix *ip-* would be used with an animate plural controller, or with a noun from Gender IV. However, gender of verbal arguments triggers more dramatic alternations elsewhere in the verb, as we will now see.

I refer to bound person markers on the verb as participant indexing since they express person/number of participants of the verb directly – there is no need to say that the affixes in (7) ‘agree’ with some ellipsed or covert argument in the clause.

- (7) *no- y-amuk-e*  
 1.A- 2SG.U-kill-IPFV  
 ‘I’m going to kill you.’

There are also frequent mismatches ('disagreement') within person indexing of a type that is not found in the gender agreement. For example, many intransitive verbs use a suppletive stem with plural subjects, with the additional quirk that actor indexing then is obligatorily 3SG instead of 3PL. Compare the regular verb *dahetok* 'return', which employs the expected 3PL indexing, with the suppletive stem *nayam* 'come (plural subject)' (cf. *man* 'come (singular subject)').

- (8) *na- dahetok*  
 3PL.A- return  
 'They returned.'

- (9) *a- nayam*  
 3SG.A- many.come  
 'They came.'

For this reason I prefer to maintain a terminological distinction between agreement and indexing in the description of Coastal Marind. I use agreement about the prefixes whose shape reflect gender and which apparently derive from relatively recently incorporated pronominal elements, while indexing is used for the markers that primarily code person/number of various argument roles, and often require construction- or verb-specific rules for their description (as in the case with the suppletive verbs above). Having established this, we are now ready to explore how gender is manifested in person indexing on the verb.

Let us start by the indexing of undergoer participants. Since we will be concerned with the difference between animate and inanimate undergoers, the discussion will be restricted to 3rd person forms (1st and 2nd person are always animate). Undergoer indexing is realized by means of intricate changes in the verb stem, and is mainly pre-, in-, or suffixing depending on the conjugation class. I will not attempt to segment the verb stems in the interlinear examples below into morphemes; the morphological details are not of interest here.

Consider the verb 'put on a string', which has the following forms when the undergoer is animate:

- (10) a. *awe ah- laleh!*  
 fish(II) IMP- string:3SG.U  
 'String one fish!'  
 b. *awe ah- lalah!*  
 fish(II) IMP- string:3PL.U  
 'String many fish!'

With inanimates from Gender III, a different stem *lalog* is used (11). Recall that no number distinction is made for inanimates, so *lalog* can be used for one or several pieces of meat, fruits, or other inanimate entities as long as they are in Gender III.

- (11) *muy ah- lalog!*  
 meat(III) IMP- string.inanimate  
 ‘String the piece(s) of meat!’

With undergoers from Gender IV, however, the stem used with animate plurals, i.e. the 3PL stem *lalah*, is used (12). As in the previous example, there is no number distinction, so the cardinality of *baba* (a kind of grass, seeds of which are used for necklaces) has to be inferred from context.

- (12) *baba ah- lalah!*  
 Job’s Tears(IV) IMP- string:3PL.U  
 ‘String the *baba* seed(s)!’

It is remarkable that Gender IV nouns trigger the use of verb stems otherwise used for 3rd person animate plurals, since gender agreement is not manifested elsewhere in person indexing. No distinction is made between Gender I and II, and inanimate stems such as *lalog* generally look like separate lexemes rather than inflectional forms of the verb. Some more examples of alternations are given in (13).

- (13) Stem alternations according to undergoer

a. ‘wrap’				
Animate	3SG:	<i>ambeh</i>	3PL:	<i>ambah</i>
Inanimate	III:	<i>ambam</i>	IV:	<i>ambah</i>
b. ‘rub (bodypart)’				
Animate	3SG:	<i>hwahwetok</i>	3PL:	<i>hwahwituk</i>
Inanimate	III:	<i>hwahwid</i>	IV:	<i>hwahwituk</i>
c. ‘eat’				
Animate	3SG:	<i>aheb</i>	3PL:	<i>hi</i>
Inanimate	III:	<i>yi</i>	IV:	<i>hi</i>
d. ‘become’				
Animate	3SG:	<i>win</i>	3PL:	<i>in</i>
Inanimate	III:	<i>ay</i>	IV:	<i>in</i>

Such verbs differ in the degree of similarity between the different stems, but all employ the same stem for Gender IV undergoers as for 3PL animates. There seem to be no exceptions to this pattern, so if a verb is semantically compatible with both animates and inanimates, then the 3PL/IV stem sharing occurs, regardless of how the remainder of the paradigm is structured. Note also that there is no morphological resemblance to the agreement patterns that we observed for nominals: with the exception of stems like *hwahwituk* ‘rub many animates’ (e.g. when scaling fish) or ‘rub a Gender IV-item’ (e.g. a knee, *mig*), which shows the high vowels /i u/ associated with gender agreement (e.g. *ihu* ‘ripe:IV’), the vowel alternations seen within the nominal domain are absent. I take this to confirm that gender agreement and participant indexing are two quite distinct phenomena in Coastal Marind, and that they have different histories, which renders the conflation of animate 3PL and Gender IV across the two systems the more remarkable.

Finally, let us consider other types of participant indexing on the verb. There are three varieties of indexing, all realized by prefixes, in addition to the indexing of undergoers by means of stem alternations. These are indexing of actor, seen in examples (7)–(9) above, plus indexing of a recipient-like participant, and what can be described as affected possessor of an argument of the verb. I will not provide examples of the latter two, because inanimate arguments filling recipient- and possessor-like roles are extremely rare in the corpus, and it is not clear whether these indexing mechanisms interact with the gender membership of inanimate arguments. The data from actor indexing are more interesting, so let us have a look at it to see whether Gender IV nouns trigger 3PL indexing in this domain.

Sentences with inanimate nouns functioning as semantic agents are also exceedingly rare in my corpus, since argument NPs headed by such nouns mostly fill patient-like roles. I have made several attempts to elicit sentences in which various things belonging to Gender IV are in violent contact with an animate undergoer (such as fruit falling from a tree, hitting a bystander), i.e. verbs that usually provide a good frame for testing all person/number combinations of agent and patient. Speakers were consistent in reporting that only 3SG actor indexing is compatible with IV agents, as in (14).

- (14) *saley*                      *a-*            *n-asib*  
       inflorescence(IV) 3SG.A- 1.U-hit  
       ‘The coconut inflorescence (fell and) hit me.’

If this were the whole story, agent indexing would finally provide an environment where Gender IV nouns were distinguished from animate plurals. However,

the generalization only seems to hold for the transitive agent-patient configuration: a small number of examples of agentive intransitives in my corpus, such as *esol* ‘make noise’ (15), unambiguously show 3PL actor indexing IV nouns (this has also been confirmed in elicitation).

- (15) *yaba-mesin i-pe t-i-k-at-n-esol-e*  
 big-machine(IV) IV-that GIV-IV-PRS-PRSTL-3PL.A- make.noise-IPFV  
 ‘The generator is making noise.’

Not even actor indexing is immune to the IV-as-animate-plural pattern, then. I take the difference in indexing between (14) and (15) to reflect semantic restrictions on what participants may be indexed on the verb, so that the inanimate coconut inflorescence in (14) is not enough of an agent to be properly indexed (with actor indexing then defaulting to 3SG, which is also the default for avalent verbs). The verb *esol* ‘make noise’ is less picky and admits its sole argument to be fully indexed, thus giving the 3PL prefix. (Recall that agreement is insensitive to number of inanimates, which means that ex. (15) is equally fine referring to one or more than one generator.)

Whatever the explanations for the subtleties of person indexing turn out to be, the data presented above are roughly consistent with the main point of this and the previous section: in all contexts where Coastal Marind, by various grammatical means, distinguishes between gender, number and animacy, nouns of Gender IV systematically pattern with plurals of Gender I and II. This is quite strange given the fact that inanimates do not show grammatical agreement according to their referential cardinality in the language (cf. example (3) above), which makes it difficult to claim that Gender IV should be considered ‘fixed plural’ nouns (*pluralia tantum*) instead of a gender. Below I will show that some tendencies in the assignment to Gender IV also are consistent with the *pluralia tantum* analysis, because they involve nouns that are *pluralia tantum* cross-linguistically. However, I will argue that this can at most be regarded as suggesting a diachronic relationship with *pluralia tantum* nouns, and that synchronically we must reject the description of the Gender IV nouns as *pluralia tantum* (§6).

## 5 Assignment and *pluralia tantum* as a possible origin for Gender IV

The basic principles behind the assignment of nouns to the four genders were given above: male humans are Gender I, female humans and all animals are Gender II, while inanimates are mostly in Gender III with a (large) residue in Gender

IV. I do not believe that there are any clear semantic rules for deciding which of the inanimates go into Gender IV, but there are some tendencies. The only semantic fields that are completely restricted to Gender III seem to be abstracts (e.g. *mayan* ‘language, issue, problem’, *sal* ‘taboo’), names of places and geographical features (*milah* ‘village’, *mamuy* ‘savannah’), and various intangibles (*matul* ‘shade’, *usus* ‘afternoon’). Other large semantic fields such as bodyparts and flora are split between Gender III and IV, with very few obvious subdomains assigned to one or the other (flowers is a subdomain that seems to belong to Gender IV). Artifacts are also divided between III and IV, with the only discernible patterns being that almost all bodily decorations are in Gender IV (*segos* ‘rattan girdle’, *himbu* ‘feathered hairdress’), as well as most recently introduced technology (airplanes, ballpoint pens, diesel generators).

Looking closer, we can see that some of the domains that Koptjevskaja-Tamm & Wälchli (2001: 630) identify as typically including pluralia tantum nouns show overlap with the members of Gender IV. These domains are: VARIOUS HETEROGENEOUS SUBSTANCES (“with many subdivisions”, e.g. Lithuanian *putos* ‘foam’), corresponding to Coastal Marind IV nouns such as *ndalom* ‘foam’, *ndakindaki* ‘bioluminescence’, *kangging* ‘layer of crushed seashells on the beach’ and *katal* ‘money’<sup>7</sup>; ARTIFICIAL OBJECTS WHICH ARE CLEARLY INTERNALLY COMPLEX (e.g. English *trousers*), corresponding to Coastal Marind decorations and modern technology in Gender IV; DISEASES “[that] manifest themselves as multiple visible symptoms/spots” (e.g. English *measles*), corresponding to names of skin diseases in Coastal Marind, which all turn out to be in Gender IV, such as *kambi* ‘tinea imbricata’, *dapadap* ‘tinea versicolor’ and *apupin* ‘pimple’.

While suggestive, these findings do not form any consistent pattern. The overlap is not found with other pluralia tantum domains such as names of festivities in Coastal Marind (e.g. German *Weihnachten* ‘Christmas’), and there are numerous exceptions, e.g. some artifacts that clearly qualify as internally complex (e.g. *kipa* ‘net’) are in Gender III rather than IV. It is also clear that – even allowing for some semantic latitude – the majority of nouns in Gender IV do not fit into any of Koptjevskaja-Tamm and Wälchli’s categories. I have found no reason why some names of trees are in Gender III, others in Gender IV, and it seems unlikely that plurality should have anything to do with the classification. Similarly, while it is conceivable that many bodyparts in Gender IV are somehow ‘plural’ (e.g.

---

<sup>7</sup>The noun *katal* has a primary use as a Gender III noun, then with the meaning ‘stone’. South New Guinea is almost completely devoid of stones, and it is extremely unlikely that one encounters two or more naturally occurring stones at the same occasion. The Gender IV noun ‘money’, on the other hand, usually occurs in collections of more than one rupiah banknote. This is an interesting case of cross-classification seemingly involving a difference in plurality.

*put* ‘feather’, *tatih* ‘hair’, *tiwna* ‘gums’, *halahil* ‘lungs’) there are plenty that are not (*ambay* ‘uvula’) and some bodyparts seem quite plural but belong to Gender III (*lul* ‘fur’). As pointed out by an anonymous reviewer, however, most languages with pluralia tantum have a fairly idiosyncratic assignment to the class, so the lack of consistency can hardly be an argument *against* the possibility of Gender IV being related to pluralia tantum.

If we consider there to be at least some tendency for ‘pluralia tantum concepts’ to be in Gender IV, this situation could be seen as consistent with a diachronic scenario where Gender IV started out as a class of pluralia tantum, but then acquired new members through some unknown (analogical?) process, resulting in a large, semantically heterogeneous residue gender, with a small core that still reflects the ‘plural semantics’ of the original pluralia tantum grouping. This scenario is only plausible if (pre-)proto-Anim (as-opposed to present-day Coastal Marind) had a number distinction among inanimate nouns, since this would be required for inanimate pluralia tantum nouns to come into existence. Also, we would expect to find some other Anim language that has been more conservative in this regard, and maintains a clearer semantically plural basis for the cognate fourth gender. Unfortunately, there is no systematic data on gender available from other Anim languages to see whether such semantics can be associated with Gender IV, nor is there any indication that proto-Anim had a number distinction among inanimates. For now this hypothesis remains purely speculative, and it can only be evaluated once there is more data on gender systems in other sub-branches of Anim. Still, I believe it is worth spelling out this hypothesis, since it has the merit of providing an explanation to the recurrent pattern of homophony between Gender IV and animate plurals, as well as the surprising phenomenon of the suppletive plural stems triggered by all Gender IV nouns.

Interestingly, a striking parallel to the Coastal Marind case is found in the Ok family, located in the New Guinean highlands. The Ok languages are probably very distant relatives of Coastal Marind and the other Anim languages as both families are proposed members of the large Trans-New Guinea phylum (Fedden 2011; Usher & Suter 2015). I believe that the Ok data support the idea that the similarities between the fourth gender of Coastal Marind (and other Anim languages) and what is described as pluralia tantum nouns in other languages are not coincidental, and perhaps that a diachronic relationship between these categories is plausible.

The best described Ok language, Mian, has a 4-gender system distinguishing Masculine, Feminine, and two inanimate genders – this is the same division as

in the gender systems of the Anim languages.<sup>8</sup> The exponents of Masculine and Feminine resemble the ones found on demonstratives in Coastal Marind (Fedden 2011: 170, Usher & Suter 2015: 118): the Mian Masculine article =*e*, the Feminine =*o*, and M/F plural =*i* correspond to Coastal Marind Gender I *epe*, Gender II *upe* and Gender I/II plural *ipe* respectively. The phonological similarities might be due to chance, however, and I am not aware of any other evidence that the gender systems of the two families are cognate. Neuter 1 (the third gender) differs from the Coastal Marind inanimates in distinguishing singular and plural (SG =*e*, PL =*o*). The most interesting gender is the fourth (“Neuter 2”) which is invariant for number, and shows homophony with the plural of Neuter 1 (SG/PL article =*o*).

It is interesting that both Coastal Marind and Mian have one gender that shares their exponents with plurals, but note that the pattern of syncretism is different (homophony with inanimate plural in Mian, but with animate plural in Coastal Marind), and could have arisen by chance since both languages have relatively few vowels to choose from (5 in Coastal Marind, 6 in Mian). Speaking against accidental homophony is the fact that even in cases where several paradigm slots are filled by unpredictable gender exponents, Neuter 2 invariably patterns with the plural of Neuter 1 (Fedden 2011: 178–179).

A further argument against the possibility of chance homophony between the Mian Neuter 2 and the plural of Neuter 1 is the fact that the nouns that are assigned to Neuter 2 match the pluralia tantum domains listed by Koptjevskaja-Tamm and Wälchli quite well – better than the Coastal Marind Gender IV nouns do. Assigned to Mian Neuter 2 we find: places (e.g. *bib* ‘village, place’), heterogeneous substances (e.g. *difb* ‘rubbish’, *moni* ‘money’), body decoration (e.g. *amún* ‘hole in nosetip’), various abstracts and temporal nouns (e.g. *am* ‘day’), illnesses (e.g. *klō* ‘ringworm’), various artifacts (e.g. *itō* ‘tongs’, *aiglas* ‘glasses’) and bodyparts, most of which seem to consist of multiple parts (e.g. *abó* ‘testicles’, *amuntēm* ‘intestines, belly’, *wanáan* ‘feather’).<sup>9</sup>

Fedden does not consider the alternative analysis according to which the Neuter 2 nouns are pluralia tantum nouns belonging to Neuter 1, and I will not pursue that issue here.<sup>10</sup> However, I interpret the parallelism between Coastal

<sup>8</sup>Sebastian Fedden (pers. comm.) adds the caveat that little is known about the gender systems of other Ok languages, so we do not know how representative the Mian system is for Ok in general. More descriptive work will be necessary for a fuller picture of the similarities and differences between the Anim and Ok gender systems.

<sup>9</sup>One instance of cross-classification is striking: Mian *bém* ‘worm’ (masculine gender) can also mean ‘noodles’, and then belongs to Neuter 2; cf. Coastal Marind *alalin* ‘tapeworm’ (Gender II), meaning ‘noodles’ in Gender IV.

<sup>10</sup>The reader is referred to Corbett et al. (2017).



Marind Gender IV and Mian Neuter 2 as further evidence that the connection between fixed plural and fourth gender in Coastal Marind is no coincidence, as this pattern would not arise independently in the two languages by chance. At this stage it is impossible to tell why the gender systems of Ok and Anim share these similarities. The two families are most likely related as members of the Trans-New Guinea stock, but this relationship is extremely distant and must go back long in time. There is at present no evidence that the gender systems were inherited from some common ancestor, although this would account for the similarities in the gender exponents mentioned above. One could also speculate that the gender systems evolved in parallel at a time when speakers of Ok and Anim languages were in closer contact, but more research remains to be done before we can say anything about the contact between these ancestral populations.

Regardless of whether the similarities between Ok and Anim are the result of common inheritance or contact, it seems to me that the simplest explanation is that both the Anim fourth gender and the Mian Neuter 2 developed from pluralia tantum nouns, which explains e.g. the use of suppletive agreement targets in Coastal Marind and the fact that many of the Mian Neuter 2 nouns (and some of the Gender IV nouns in Coastal Marind) have meanings that are found among pluralia tantum cross-linguistically. This hypothesis can be tested only through more descriptive and comparative work on the two families. Even if it is correct, it would still remain to be shown in detail how a 3-gender system with a large number of pluralia tantum nouns can develop into a 4-gender system lacking number distinction in inanimates, as in present-day Coastal Marind.

## 6 The synchronic analysis of Gender IV

Having suggested that the Coastal Marind Gender IV originated as a pluralia tantum class, we now need to address the synchronic status of Gender IV. Should we maintain the 4-gender analysis, or opt for the more economical 3-gender analysis according to which the members of the former fourth gender are Gender I or II nouns that just happen to be lexically specified as plural? I believe that this is an important analytical question – not a mere question of which labels to stick where – since the two possible descriptions result in wildly different systems in terms of assignment.

The literature contains some discussion of the possibility of analyzing pluralia tantum as a separate gender, in various languages. Corbett (2012: 233–239) provides instructive discussion of such suggestions for Cushitic, Chadic and Russian, and argues that the pluralia-tantum-as-gender analysis is untenable for all

the proposed cases (i.e., the opposite of the established descriptions of Coastal Marind and Mian). For example, Zaliznjak (1964) proposed to describe Russian pluralia tantum nouns such as *sani* ‘sledge(s)’ as making up their own gender, since they form a unique agreement class within the system. Corbett (2012: 237–238) points out that the same analysis applied to Bosnian/Croatian/Serbian would produce no less than three extra genders, since this three-gender system (as opposed to Russian) has separate plural forms for each gender, each of which contains pluralia tantum that would be reanalyzed as separate genders. This is unacceptable, so Corbett rejects the analysis for Russian as well.

On a more general level, Corbett argues that pluralia-tantum-as-gender analyses are misinformed, since “the special behaviour which creates the extra agreement class is not *gender* but *number*” (Corbett 2012: 238; emphasis in original). According to Corbett, proponents of pluralia-tantum-as-gender analyses mistakenly think that since pluralia tantum nouns need to be lexically specified for a morphosyntactic value (in this case number), they are just like other nouns – which are also lexically specified, for gender – and therefore belong to a gender of their own. Instead, the correct way is to treat them as exceptionally specified for number, and leave the gender system as it is. I interpret Corbett’s remarks as a principled stance against analyses claiming that pluralia tantum nouns make up a gender.<sup>11</sup>

In spite of Corbett’s reservations, I prefer to maintain the Drabbian analysis of Gender IV as a gender, and not as pluralia tantum of Gender I or II, although I concede that the morphosyntactic evidence for this analysis is somewhat nebulous. We saw that the exponents of Gender IV agreement are identical to the ones marking the plural of Gender I and II, no matter how irregular the alternations of the relevant target are. Verb stem alternations indexing undergoers likewise treat Gender IV and plurals of I/II identically, despite being seemingly unrelated to the agreement patterns of demonstratives and other categories in the non-verbal domains. The only domain where Gender IV nouns do not always pattern with I/II plural is actor indexing (and, possibly, recipient and possessor indexing) on verbs; however, I suspect that this reflects some general constraint against inanimates filling such participants roles, so the diagnostic role of these constructions is unclear.

But consider the consequences of abandoning the gender analysis in favour of the pluralia tantum analysis. If the members of Gender IV are considered plu-

---

<sup>11</sup>In fact, Corbett says explicitly that this is what he means: “Having not accepted Zaliznjak’s careful and considered analysis of certain Russian pluralia tantum nouns as an additional gender value, I am even less ready to entertain other less convincing proposals along similar lines.” (p. 238).

ralia tantum, they would make up an unexpectedly large portion of the lexicon. Assuming that the currently available numbers (Table 3) are representative of gender membership, one out of five nouns would be pluralia tantum. This seems strange from the European perspective, but sheer frequency can hardly be a decisive argument. More seriously, the system of semantic assignment (males in I, females and animals in II, inanimates in III and IV) would break down, since we would have to claim that Gender I and II contain a fairly random mix of animates and inanimates (all of which happen to be pluralia tantum), with non-pluralia tantum inanimates confined to Gender III.

The resulting system would also be typologically odd in the way it fails to align with the Animacy Hierarchy (Smith-Stark 1974, Corbett 2000: 55ff.). The hierarchy states that if there is a difference in the availability of a number distinction between e.g. animates and inanimates, then it will be animates that make the distinction and inanimates that lack it. Corbett (2000: 59) cites Coastal Marind as an example of a language with a clear split between animates (which trigger singular/plural agreement) and inanimates (which make no distinction according to number). In the new system, we would have to say that number is relevant for a fifth of the inanimates, although these happen to be lexically specified for plural only.

I take these consequences to be unacceptable, so the 4-gender analysis must be preferred. This comes at the price of not adhering to a strictly morphosyntactic approach to the identification of genders in Coastal Marind, because the formal facts alone do not provide clear evidence that the four-gender description is to be preferred over a three-gender description with a large number of pluralia tantum.

## 7 Conclusion

Besides the descriptive contribution of this paper (most of which can be extracted, with some effort, from Drabbe's grammar), I consider the main points to be (1) the evidence that Usher & Suter's (2015) suggestion that overt, stem-internal gender marking originated from umlaut also explains patterns in the distribution of stem-final vowels of invariant nouns within Gender III and IV; and (2) the description of the ambiguous status of the nouns in Gender IV, which led me to speculate that an earlier 3-gender system was extended into a 4-gender system, and that the 4th gender originally was a grouping of pluralia tantum nouns. As mentioned above, the idea that gender systems can be extended through the reinterpretation of a non-gender feature as gender is not new, and if the suggestions based on Coastal Marind data are correct, the Anim languages (and the distantly

related Ok family) would provide a clear case where a gender system became more complex because of a very specific type of interaction with number.

## Acknowledgments

I am very grateful to the Coastal Marind speakers with whom I work, especially Petrus Kilub and Rafael Samkakai who spent many hours with me rechecking the gender of nouns. I wish to thank Matthew Lou-Magnuson for suggesting the use of resampling methods in §2 and Thomas Hörberg for implementation; I alone am responsible for mistakes in the interpretation of the data. I also acknowledge the extremely helpful comments given to me by Edgar Suter, Sebastian Fedden, Bernhard Wälchli, Francesca Di Garbo, Lea Brown, and the anonymous reviewers.

## Special abbreviations

The following abbreviations are not found in the Leipzig Glossing Rules:

A	actor	PRSTL	presentational
ABSC	absconditive	U	undergoer
GIV	givenness marker		

## References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Brugmann, Karl. 1891. Zur Frage der Entstehung des grammatischen Geschlechts. *Beiträge zur Geschichte der Deutschen Sprache und Literatur* 15. 523–531.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2012. *Features*. Cambridge: Cambridge University Press.
- Corbett, Greville G., Sebastian Fedden & Raphael Finkel. 2017. Single versus concurrent feature systems: Nominal classification in Mian. *Linguistic Typology* 21(2). 209–260.
- Dahl, Östen. 2000. Animacy and the notion of semantic gender. In Barbara Unterbeck (ed.), *Gender in grammar and cognition*. Vol. 1: *Animacy and the notion of semantic gender: Approaches to gender*, 99–115. Berlin: Mouton de Gruyter.
- Drabbe, Peter. 1955. *Spraakkunst van het Marind: Zuitkust Nederlands Nieuw-Guinea* (Studia Instituti Anthropos 11). Wien-Mödling: St. Gabriël.

- Durie, Mark. 1986. The grammaticization of number as a verbal category. In Vasiliki Nikiforidou, Mary VanClay, Mary Niepokuj & Deborah Feder (eds.), *Proceedings of the 12th annual meeting of the Berkeley Linguistics Society*, 355–370. Berkeley: Berkeley Linguistics Society.
- Edwards-Fumey, Deborah. 2007. *The verb subject prefix in Kuni*. Universität Bern MA thesis.
- Evans, Nicholas. 2012. Even more diverse than we had thought: The multiplicity of Trans-Fly languages. In Nicholas Evans & Marian Klammer (eds.), *Melanesian languages on the edge of Asia: Challenges for the 21st century* (Language Documentation & Conservation Special Publication No. 5). Honolulu: University of Hawai'i Press.
- Evans, Nicholas, Wayan Arka, Matthew Carroll, Yun Jun Choi, Christian Döhler, Volker Gast, Eri Kashima, Emil Mittag, Bruno Olsson, Kyla Quinn, Dineke Schokkin, Philip Tama, Charlotte van Tongeren & Jeff Siegel. 2018. The languages of Southern New Guinea. In Bill Palmer (ed.), *The languages and linguistics of the New Guinea area* (The World of Linguistics 4), 641–774. Berlin: De Gruyter Mouton.
- Fedden, Sebastian. 2011. *A grammar of Mian*. Berlin: Mouton de Gruyter.
- Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- Geurtjens, Hendrik. 1933. *Marindineesch-Nederlandsch woordenboek* (Verhandelingen van het Koninklijk Bataviaasch Genootschap van Kunsten en Wetenschappen 71). Bandung: Nix.
- Koptjevskaja-Tamm, Maria & Bernhard Wälchli. 2001. The Circum-Baltic languages: An areal-typological approach. In Östen Dahl & Maria Koptjevskaja-Tamm (eds.), *Circum-Baltic languages: Typology and contact*, vol. 2, 615–750. Amsterdam: John Benjamins.
- Olsson, Bruno. 2017. *The Coastal Marind language*. Singapore: Nanyang Technological University dissertation.
- Smith-Stark, Thomas Cedric. 1974. The plurality split. In Michael W. La Galy, Robert A. Fox & Anthony Bruck (eds.), *Papers from the 10th regional meeting of the Chicago Linguistic Society*, 657–661. Chicago: Chicago Linguistics Society.
- Usher, Timothy & Edgar Suter. 2015. The Anim languages of Southern New Guinea. *Oceanic Linguistics* 54(1). 110–142.
- Zaliznjak, Andrej A. 1964. K voprosu o grammatičeskix kategorijax roda i oduševlennosti v sovremennom russkom jazyke. *Voprosy jazykoznanija* 4. 25–40.



## Chapter 9

# Gender in New Guinea

Erik Svärd

earlier Stockholm University

The present study classifies gender systems of 20 languages in the New Guinea region, an often neglected area in typological research, according to five criteria used by Di Garbo (2014) for African languages. The results show that gender in New Guinea is diverse, although around half of the languages have two-gendered sex-based systems with semantic assignment, more than four gender-indexing targets, and no gender marking on nouns. The gender systems of New Guinea are remarkably representative of the world, although formal assignment is underrepresented. However, the gender systems of New Guinea and Africa are very different. The most significant difference is the prevalence of non-sex-based gender systems and gender marking on nouns in Africa, whereas the opposite is true in New Guinea. Finally, four typologically rare characteristics are singled out: (1) size and shape as important criteria of gender assignment, with large/long being masculine and small/short feminine, (2) the co-existence of two separate nominal classification systems, (3) no gender distinctions in pronouns, and (4) verbs as the most common indexing target.

**Keywords:** agreement, grammatical gender, indexation, New Guinea, Papuan, typology.

## 1 Introduction and background

Most typological research on gender has focused on languages in Eurasia, Africa, Australia, and the Americas. Less research has been conducted in the region of New Guinea, which contains as many as one sixth of all languages of the world. In recent descriptions, languages of New Guinea of highly variable genealogical affiliation have been shown to exhibit many unusual gender systems. This is important for the study of gender as gender systems are often very stable and not prone to borrowing. However, little has been done to survey the diversity of



gender in New Guinea. The purpose of this paper is to counteract this issue by investigating 20 New Guinean languages, both Papuan and non-Papuan, for which gender has been described and to compare their gender systems in an areal and a typological perspective. Specifically, the research questions are:

- How is grammatical gender expressed in a diverse sample of 20 languages of New Guinea?
- How do the gender systems of New Guinea compare with other geographical areas (notably Africa) and the world as a whole?
- Are there any phenomena in gender which are unique to or surprisingly common in the languages of New Guinea?

In order to investigate this, five criteria are used to classify the gender systems of New Guinea. The distribution of values of these criteria are then compared with the world in general and Africa in particular.

## 1.1 Defining gender

Hockett (1958: 231) defines gender as “classes of nouns reflected in the behavior of associated words”. In other words, gender is conceived of as noun classes triggering agreement. The idea of gender as based on the behavior of associated words is reflected in the focus on agreement, which Corbett (1991: 4) calls the determining criterion of gender. In order to define gender, Corbett presents Steele’s (1978) description of agreement:

The term *agreement* commonly refers to some systematic covariance between a semantic or formal property of one element and a formal property of another. For example, adjectives may take some formal indication of the number and gender of the noun they modify.  
(Steele 1978: 610 as cited in Corbett 1991: 105)

According to Corbett, agreement is an asymmetric relationship between the controller (i.e., the element determining agreement, e.g., subject noun phrase) and the target (i.e., the element whose form is determined by agreement) (Corbett 2006: 4). Importantly, Corbett adopts a ‘canonical approach’: that is, the basis for Corbett’s discussion are those ‘canonical’ instances which are best and clearest but not necessarily the most frequent (Corbett 2006: 9). Canonical agreement can be summarized as follows (adapted from Corbett 2006: 9):



- the controller is present, has overt expressions of features, and is consistent in agreement, and its part of speech is not relevant;
- the target has bound expression of agreement, obligatory and regular marking which is doubling the marking of the noun, has a single controller, and its part of speech is not relevant;
- the domain in which agreement occurs is local, and it is one of multiple domains.

More recently, Di Garbo (2014: 8) gives a few examples illustrating the fact that in many languages both pronouns and noun phrase-internal targets do not presuppose a syntactic antecedent or controller. In order to counter this, Di Garbo (2014) uses the term *indexation* instead, following Croft (2001; 2003; 2013) and Iemmolo (2011). In this definition, indexation is used to refer to grammatical strategies signaling (i) lexical and grammatical properties of nouns, and (ii) semantic properties of NP referents, which are independent of the presence of any overt syntactic antecedent (Di Garbo 2014: 8). Following Di Garbo, the following terms are used in this study (adapted from Di Garbo 2014: 8):

- *indexing target* or *index* refers to entities with inflectional morphology signaling gender;
- *syntactic antecedent* refers to the NP indexed by the pronominal target;
- *indexation trigger* or *trigger* refers to the entities that activate the use of a certain indexation pattern in a given discourse domain.

Despite the difference in terminology, the end result of both agreement in Corbett (1991) and indexation in Di Garbo (2014) is the same, with both being cover terms for the same linguistic feature. Since this is mainly a typological study, its purpose is to be comparable with earlier and future typological research on gender without relying on theoretical concepts that are as yet not widely accepted. However, since indexation is gaining ground, it is embraced in this chapter.

## 1.2 Gender research on New Guinea

Although gender has not been extensively researched in New Guinea, the region shows much promise for exhibiting a high variety of gender systems. The New Guinea region is home to approximately 1,200 languages belonging to around three dozen language families spoken in an area smaller than 900,000 km<sup>2</sup>, which

makes it the most linguistically diverse region in the world (Foley 2000: 357). Nevertheless, there are two dominating language families: the Austronesian family, spoken in the coastal areas, and the Trans-New Guinean (TNG) family, which is concentrated to the mountainous inland. The Austronesian and the TNG languages comprise around 300 languages each and typically do not show gender, although there are some important exceptions (Foley 2000: 358–363). Thus, gender is lacking at least in approximately half of the languages of New Guinea. As for the remaining languages, gender is found in the West Papuan, Sko, and Sepik languages, as well as several isolates such as Yava, Burmeso, and Kuot (Foley 2000: 371).<sup>1</sup> Gender is also present in Torricelli and Lower Sepik-Ramu languages, but as parts of larger and more complex systems of noun classification (Foley 2000: 371). It also occurs in some isolated cases in the TNG family, such as Nalca (Mek) (Svärd 2013) and the Ok languages, e.g., Mian (Fedden 2011), and in very few Austronesian languages, including Teop (Oceanic) (Mosel & Spriggs 2000). By counting these gendered languages based on the numbers given by Foley (2000), gender in New Guinea can be estimated to occur in at least 120 languages of different families and isolates. The genealogical diversity suggests that gender may be highly diverse in New Guinea.

However, Foley suggests that gendered languages of New Guinea have some features in common, including the presence of gender assignment based on specific criteria of size and shape, as well as the presence of languages with two separate systems (Foley 2000; Svärd 2015: 8–9). Combined with the observation that gender in New Guinea is concentrated in languages with high genealogical diversity, this suggests that gender may be highly diverse in New Guinea.

## 2 Method and data

The sampling method used in this study is a *variety sample* (Bakker 2012). Rather than trying to represent the real population of languages as would be achieved by a probability sample, the sample is designed to achieve the largest variety of results in regard to the chosen feature, while entirely omitting languages lacking the feature.

In this study, the sample is restricted to New Guinea as delimited by Foley (2000: 357), including New Guinea proper as well as surrounding islands. First and foremost, the sample includes only languages with gender. Secondly, the languages were chosen from as many families as possible, as far as the availability

---

<sup>1</sup>Foley (2000: 371) also mentions the Sulka language of New Britain, but there are no indications of a gender system in the grammar by Tharp (1996).

of material permitted, while still accounting for variation within families if there were reasons to do so. This was primarily based on the information by Foley (2000) and others.

Table 1 lists the languages of the sample together with family, genus, ISO code, and source, along with a map of the languages shown in Figure 1. The names primarily follow Glottolog, except for Motuna (Glottolog: Siwai), where I follow Onishi (1994). Also, the Glottolog form Warapu is used despite Barupu occurring in Corris (2005). Furthermore, language families and genera are based on Glottolog, so that a genus in the table below does not always agree with the genus for the same language in *WALS*.

Table 1: The language sample. The en dash indicates no grouping or that the language is itself the closest node to the family node.

Family	Genus	ISO	Language	Source
Austronesian, Oceanic	Nehan-North Bougainville	tio	Teop	Mosel & Spriggs (2000)
Isolate	–	gpn	Taiap	Kulick & Stroud (1992)
Isolate	–	bzu	Burmeso	Donohue (2001)
Isolate	–	kto	Kuot	Lindström (2002)
Left May	–	amm	Ama	Årsjö (1999)
Lower Sepik-Ramu	Lower Sepik	yee	Yimas	Foley (1991)
Ndu	–	mle	Manambu	Aikhenvald (2008)
North Bougainville	–	roo	Rotokas	Robinson (2011)
Sepik	–	aau	Abau	Lock (2011)
		sim	Mende	Hoel et al. (1994)
Sko	–	skv	Skou	Donohue (2004)
		wra	Warapu/Barupu	Corris (2005)
South Bougainville	–	siw	Motuna/Siwai	Onishi (1994)
Torricelli	–	avt	Au	Scorza (1985)
	Arapesh	ape	Bukiyip	Conrad & Wogiga (1991)
	West Palai	van	Walman	Brown & Dryer (2008)
Trans-New Guinea	Mek	nlc	Nalca	Svärd (2013); Wälchli (2018)
	Ok-Oksapmin	mpt	Mian	Fedden (2011)
		opm	Oksapmin	Loughnane (2009)
West Papuan <sup>2</sup>	–	ayz	Maybrat	Dol (2007)



Figure 1: The geographical locations of the languages in the sample labeled with ISO codes

The main sources of data used in this study are reference grammars, which are listed for each language in Table 1 above. However, many descriptions do not mention the language as having a gender system if gender only occurs in pronouns. Therefore it was also necessary to examine the sections on pronouns. If the available descriptions for a language neither mentioned gender nor showed it directly in the section(s) about pronouns or in glossed examples, the language was not considered to be eligible for the sample.

In order to make the languages of the study typologically comparable, the study employs five classificatory criteria used by Di Garbo (2014) to classify the gender systems of Africa, viz.,

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

Di Garbo also uses other classificatory criteria in order to investigate the interactions of gender and number, and gender and evaluative morphology. However,

---

<sup>2</sup>More recent studies suggest that the traditional West Papuan Phylum is probably not an accurate genealogical grouping, but instead consists of as many as seven unrelated language groups (Dol 2007: 5). Since the exact position of Maybrat in such a regrouping is unknown to the present author, West Papuan is kept here as proxy to a genealogical family.

this study is not aimed to directly investigate these interactions, and thus only the criteria above were chosen.

An important advantage of adopting Di Garbo's approach is that this makes the results for New Guinea directly comparable with Africa as for the selected criteria. In addition, since the first three criteria are the same as the ones used in the *WALS* chapters by Corbett (Corbett 2013a,b,c), most of the results are comparable to a worldwide sample. In order to illustrate the distributions, maps were created using the Interactive Reference Tool of the *World atlas of language structures* (*WALS*)<sup>3</sup> using ISO codes and coordinates from Glottolog.

### 3 Overview of gender characteristics

In the following sections, the distribution of values of the criteria mentioned in §3 are presented and discussed. Each criterion is discussed with the values shown in a table, followed by some examples of the feature in the sample. In §4, these results are discussed from a typological perspective.

It is important to point out that five languages of the sample were found to have two separate systems of noun classification. As will be discussed in §5.2, only Burmeso exhibits two equivalent gender systems, whereas the other four rather distinguish between gender and noun classifiers. For this reason, the two gender systems of Burmeso will be combined for the purpose of comparison in this chapter, although the values assigned to the separate systems will be given in parenthesis whenever applicable.

#### 3.1 Sex-based and non-sex-based gender systems

Following Di Garbo (2014: 62), each gender system is classified as either sex-based or non-sex-based based according to the typology by Corbett (2013c). Sex-based are those where the gender assignment is based at least partly on natural gender, which often surfaces as masculine-feminine distinctions. Consequently, non-sex-based gender systems are those where gender is not based on natural gender. However, according to Corbett (2013c), all non-sex-based systems are based on some notion of animacy.

As shown in Table 2 and Figure 2, sex-based systems are by far the most common ones, with 19 of 20 languages having natural gender as their semantic core. Only the Austronesian language Teop exhibits a non-sex-based system.

Sex-based gender systems present some difficulty in assigning nouns denoting inanimate referents. Non-sex-based systems, i.e., systems based on animacy, can

<sup>3</sup>See <http://www.eva.mpg.de/lingua/research/tool.php>.

Table 2: Sex-based and non-sex-based gender systems in the sample

Sex-based or non-sex-based	No. of lgs.	%	Languages
Sex-based	19	95%	Abau Ama Au Bukiyip Burmese Kuot Manambu Maybrat Mende Mian Motuna Nalca Oksapmin Rotokas Skou Taiap Walman Warapu Yimas
Non-sex-based	1	5%	Teop
Total:	20	100%	



Figure 2: Sex-based and non-sex-based systems. Colors indicate: sex-based (blue) and non-sex-based (red).

potentially assign every noun according to animacy alone. However, sex-based systems do not by definition have any specific way of assigning nouns that refer to objects without natural gender. Thus, based on how inanimate nouns are assigned gender, the sex-based gender systems in the sample can be further divided into three groups where inanimates are assigned to

1. one of the sex-based genders,
2. both of the sex-based genders based on other criteria, or
3. one or more other non-sex-based genders.

As will be discussed in §3.2, almost half of the languages in the sample (9 of 20) have only two genders, both of which are sex-based. Thus, since option 3 is only available in languages with more than two genders, almost half of the languages in the sample assign inanimate nouns to one of the sex-based genders.

Assigning inanimates to only one of the two genders occurs e.g., in Mende (Sepik), where gender is distinguished only in second and third person singular pronouns. For animate referents, the form of the pronoun is determined by the sex of the referent, while inanimates are usually referred to with the feminine forms (Hoel et al. 1994: 17). An example of this is shown in (1), where *Max* (male name) (1a) and *Lusi* (female name) (1b) occur with the masculine and feminine pronoun forms respectively, and the inanimate *masiji* ‘hair’ (1c) is referred to with the feminine form. Mende thus distinguishes masculine vs. other.

(1) Mende (Sepik) (Hoel et al. 1994: 19, 31, 46)

- a. *Max wasilaka ri-a*  
M. big 3SG.M-INTEN  
‘Max is big.’
- b. *Lusi kava awu-n u-nda sir-a*  
L. bad<sup>4</sup> fight-Obj do-HAB 3SG.F-INTEN  
‘Lusi is a good fighter.’
- c. *masiji-n tivi unak si horngo-ku-a*  
hair-Obj tie so.that.not 3SG.F loosen-FUT-INTEN  
‘Tie the hair so that it won’t loosen.’

<sup>4</sup>When used with the habitual *-nda*, *kava* ‘bad’ functions as an intensifier (Hoel et al. 1994: 31).

Assigning inanimates to both sex-based genders based on other criteria is more common in the sample. In most languages, the assignment of inanimates is based on semantic criteria, most commonly on the criteria shape and size (see also §5.1 below). One such a language is Abau (Sepik), where three-dimensional or long or extended objects, as well as liquids are masculine, whereas two-dimensional, flat or round objects with little height as well as abstract entities are feminine (Lock 2011: 47). Thus, *su* ‘coconut’ (round), *now* ‘tree’ (long), and *hu* ‘water’ (liquid) are masculine, while *iha* ‘hand’ (flat) and *hne* ‘bird’s nest’ (round with little height) are feminine (Lock 2011: 48–50). In a language such as Abau, this is very much based on the speaker’s perception. This can be seen in (2); when referring to the tree from which he makes the paddle (2a), *youk* ‘paddle’ is masculine, since the tree is long and not at all round or flat. However, when referring to the actual paddle (2a), which has the salient features of flat and round, the feminine form is used.

(2) Abau (Sepik) (Lock 2011: 50)

- a. *Ha-kwe youk se seyr.*  
 1SG.SBJ-TOP paddle 3SG.M.OBJ cut  
 ‘I cut the ‘paddle’ tree.’
- b. *Ha-kwe youk ke lira.*  
 1SG.SBJ-TOP paddle 3SG.F.OBJ see  
 ‘I see the paddle.’

The third type of sex-based systems is one where inanimates are assigned to genders other than sex-based ones. Naturally, this can only occur in languages with more than two genders. An example of a language with such a system is Nalca (TNG, Mek) (Svärd 2013; Wälchli 2018). Nalca has five main genders: masculine, feminine, neuter, default, and non-noun. As shown in (3), these are apparent in a set of case marker hosts following the NP, which constitute the only indexing target in Nalca. The masculine and feminine genders are used exclusively for nouns denoting male and female humans respectively. Inanimates are divided between the neuter and default genders: the neuter contains all nouns of the phonological structure (C)V (including at least one noun denoting humans, *me* ‘son, child’), while most inanimate nouns belong to the residual default gender. The default gender also contains some gender-neutral nouns denoting humans, most of which are plural, e.g., *nang* ‘people’. The non-noun gender is used e.g., with adverbs, locatives, and despite its name the nominalizer *-a*. It is also used when gender is switched off, in which case nouns still trigger agreement but



due to syntactic phenomena agree with the non-noun gender.<sup>5</sup> In the examples below, both the neuter *si* ‘name’ and the masculine name *Zakheus* ‘Zacchaeus’ are shown in (3a), the feminine *genong* ‘mother’ in (3b), the default (DEFAULT) *pik* ‘way’ in (3c), and the two non-noun (NNOUN) constructions in (3d). The first instance of non-noun gender in (3d) is due to the intervention of the quantifier *nauba* ‘many’ between *nimi* ‘men’, which belongs to the default gender, and the case marker host, whereas the second is due to the nominalizer *-a*’.

(3) Nalca (TNG, Mek) (own examples)

- a. *alja si ne-ra Zakheus be-k u-lu-m-ok*  
 3SG.GEN name N-TOP Z. M-ABS be-IPFV-PST.3SG  
 ‘a man called by name Zacchaeus’ (Lk 19:2)<sup>6</sup>  
 lit. ‘his name was Zacchaeus’
- b. *Nadya genong ge-ra heknya do?*  
 1SG.GEN mother F-TOP who Q  
 ‘Who is my mother?’ (Mk 12:48)
- c. *Na bi-nim-na pik e-ra ugun-da ella*  
 1SG go-FUT-PRS.1SG way DEFAULT-TOP 2PL-TOP knowledge  
*u-lu-lum ...*  
 be-IPFV-PRS.2PL  
 ‘And you know the way where I am going.’ (Jn 14:4)
- d. *... nimi nauba a-ra seleb longo-m-ek-a*  
 men many NNOUN-TOP already assemble-PRF-PST.3PL-NMLZ  
*a-k eib-ok*  
 NNOUN-ABS see[PFV]-PST.3SG  
 ‘... he saw the large crowds...’ (Mt 6:34; lit. ‘he saw that many men had assembled’)

Finally, the only non-sex-based gender system in the sample occurs in the Austronesian language Teop, which has two genders (I and II) with two subgenres for the first gender (I-E and I-A), reflecting the form of the singular article preceding nouns. The genders and the nouns that belong to them are:

<sup>5</sup>The concept of switching gender on and off is an extremely rare phenomenon and goes well beyond the bounds of this study. For a comprehensive description of the Nalca gender system and discussion on switching gender on and off, see Wälchli (2018).

<sup>6</sup>The overwhelming majority of data available in Nalca consists of a translation of the New Testament. The English translation used is the American Standard Version, whereas the glossings and literal translations were devised by the present author. For a description and discussion of the methodology, see Svärd (2013) and Wälchli (2018).

- *Gender I-E*: Contains all proper names, kinship terms, and nouns denoting pets or humans with a particular communal or important social status (Mosel & Spriggs 2000: 334–335).
- *Gender I-A*: Contains most nouns and can be considered the unmarked gender (Mosel & Spriggs 2000: 336–338).
- *Gender II*: Contains names of plants and their parts (but not fruits), objects made of plant material, invertebrates without legs, and many mass and abstract nouns (Mosel & Spriggs 2000: 338).

This is strikingly similar to the noun classification system found in Siar (Frowein 2011; not in the sample), spoken on the opposite coast. Siar does not have a true gender system, since it only shows gender on articles preceding nouns and thus does not exhibit indexation.<sup>7</sup> However, nouns are still assigned according to a system of nominal classification similar to Teop:

- *Proper*: Contains mostly names, kinship terms and other nouns closely related to humans and culture such as professions (Frowein 2011: 104–105).
- *Common 1*: A very heterogeneous residual class, consisting of all nouns not in the proper or common 2 genders (Frowein 2011: 108).
- *Common 2*: Contains semantically marked nouns, including entities that are smallish or individuated from a greater mass, but also other semantic types; some examples are insect, birds, other smallish animals, plants and parts of plants, tools, loanwords, geographic locations, some meteorological phenomena, groups and sets, and ordinals (Frowein 2011: 105–107).

Teop and Siar thus clearly display the differences between a gender system and a simpler noun classification system according to the criteria of gender used in this paper.

---

<sup>7</sup>In this study, a word is only considered an indexing target if it has a functional load other than expressing gender and number. The reason for this is that otherwise languages such as Siar, which has a set of markers preceding only nouns, would be considered as having gender. Such a system would be difficult to separate from a system showing noun classification only on the noun itself, i.e., without indexation.

### 3.2 Number of genders

The second criteria concerns the number of genders in a language, based on Corbett (2013a). Each language is assigned the value two, three, four, or five or more genders (see Table 3 and Figure 3). The majority of the languages have only two genders, in all cases sex-based. Only one, Mian, has four genders. Of the remaining languages, three languages have three genders, whereas the remaining five languages have five or more genders, viz., Nalca (5), Motuna (6),<sup>8</sup> Burmeso (9 [3+6]),<sup>9</sup> Yimas (around 12), and Bukiyip (18 genders).

In contrast to the previous criterion, it is more difficult to identify subgroups based on values of the number of genders; e.g., the languages with three genders are very different from each other. Nevertheless, some of the languages have the following specific characteristics of

1. two genders where one is unmarked,
2. three genders consisting of masculine, feminine, and neuter, or
3. very large systems.

More than half of the languages with two genders have one which is unmarked, all of which are sex-based. Consequently, in these languages, either the feminine or the masculine gender is unmarked. An example of such a language is Maybrat (West Papuan), which has the conveniently named genders masculine and unmarked (i.e., non-masculine) (Dol 2007: 89). Thus, nouns denoting male humans (or in some cases other male animates) are masculine, whereas all others (including those denoting females) belong to the unmarked gender. This is shown in (4). In (4a) ‘old’ indexes ‘his father’, in (4b) ‘his mother’, and in (4c) ‘big’ indexes ‘house’.

(4) Maybrat (West Papuan) (Dol 2007: 90)

- a. *y-atia*      *y-anes*  
      3M-father 3M-old  
      ‘His father is old.’/‘his old father’

<sup>8</sup>Onishi (1994) states that Motuna has six genders: masculine, feminine, diminutive, local, manner, and dual-paucal. However, the author does not elaborate on gender assignment, and I have been unable to satisfactorily conclude that the dual-paucal is truly a gender, which Onishi states. However, all form a complementary and mutually exclusive system, with separate identifiable markers and where a word may take only one gender.

<sup>9</sup>Burmeso has two gender systems, with three genders belonging to the first system and the other six belonging to the second system (see §5.2).

Table 3: Number of genders in the languages of the sample

Number of genders	No. of lgs.	%	Languages
Two	11	55%	Abau Kuat Manambu Maybrat Mende Oksapmin Skou Taiap Teop Walman Warapu
Three	3	15%	Ama Au Rotokas
Four	1	5%	Mian
Five or more	5	25%	Nalca (5) Motuna (6)
Burmeso (9 [3+6])			Yimas (~12) Bukiyip (18)
Total:	20	100%	

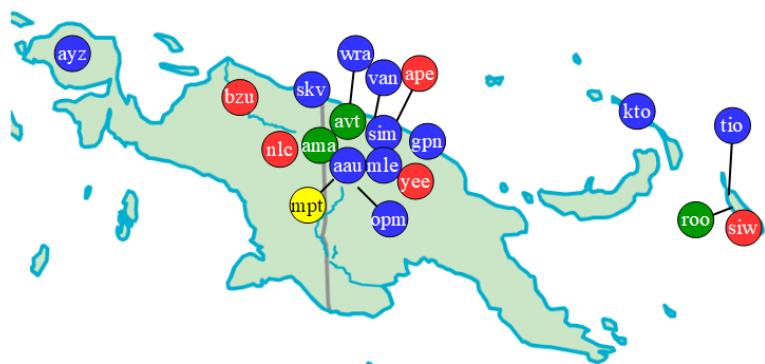


Figure 3: Number of genders. Colors indicate: two (blue), three (green), four (yellow), and five or more (red).

- b. *y-me*            *m-anes*  
 3M-mother 3U-old  
 ‘His mother is old.’/‘his old mother’
- c. *amah* *m-api*  
 house 3U-big  
 ‘The house is big.’/‘the big house’

However, not all such languages use the masculine gender as the marked one. Languages where the masculine is marked are Warapu (Sko), Maybrat (West Papuan), Mende (Sepik), and Taiap (isolate), whereas the feminine is marked in Skou (Sko). It is also marked in Ama (Left May), which has three genders: masculine, feminine, and compound. However, the situation is more complex in Ama, both because there are three genders, and because the feminine also includes e.g., some non-female animates (Årsjö 1999: 68).

Except for Ama, which is mentioned above, the three-gendered systems belong to the second type, since all have masculine, feminine, and neuter. While this implies that inanimates are found only in the neuter gender, all languages assign some inanimates to the masculine and feminine genders as well, with or without sex-based motivation. For example, in Rotokas (North Bougainville), inanimate objects associated with male culture (such as hunting or warfare) and long, thin objects are masculine (see also §5.1), whereas most inanimates are assigned either to the feminine or to the neuter genders (Robinson 2011: 46–48).

The third and final type is languages with very large gender systems, viz., Bukiyip (Torricelli, Arapesh) and Yimas (Lower Sepik-Ramu, Lower Sepik). These are markedly different from all other languages in the sample. The most immediate difference is of course the vastly larger number of genders. Bukiyip has as many as 18 genders (Conrad & Wogiga 1991: 8–10), while Yimas has around a dozen genders, with Foley (1991: 119) distinguishing 10 and Phillips (1993: 175) as many as 16. All other languages in the sample have six genders or fewer. The Bukiyip genders and their indexing forms are shown in Table 11 in §3.5. The most important feature of these two gender systems is that both have semantic-formal assignment and gender marking on nouns. These two factors, which are uncommon in the sample, are undoubtedly related to the subsistence of their large systems.

Finally, a highly interesting case is Burmeso, which is the only language in the sample with two gender systems. The first system has three genders (masculine, feminine, and neuter), each with an additional subgender for inanimates, whereas the second system has six genders (I–VI). The exact nature of the gender systems and their interaction will be discussed further in §5.2.

### 3.3 Gender assignment

The third criterion concerns gender assignment and contains two values (see Table 4 and Figure 4), viz., semantic, or semantic and formal.

As can be seen in Figure 4, the majority of languages in the sample have semantic assignment. However, there are major differences between the various semantic systems as to their complexity. As mentioned in §3.1, Mende (Sepik) has an extremely simple system of gender assignment, where all nouns denoting human or sometimes animate males are masculine while all other nouns are feminine.

In Rotokas (North Bougainville), however, the situation is more complex. Rotokas has three genders: masculine, feminine, and neuter. Both the masculine and the feminine gender contain nouns denoting male and female referents respectively, but complexity arises for inanimates. The masculine gender contains many inanimate objects, which are often associated with male culture or which are long or thin (Robinson 2011: 46). The feminine gender also contains many inanimate objects, some of which are tools or related to water, but many which have no apparent semantic or formal criteria at all (Robinson 2011: 47). Finally, as expected, many inanimate nouns belong to the neuter gender (Robinson 2011: 48).

Thus, while a learner of Mende is easily able to guess the correct gender of any noun, a learner of Rotokas is hard-pressed to guess the correct gender of an inanimate object. Even if there are rules, many of these are probably not tacitly known. Furthermore, even if the rules for gender assignment can be explicitly stated, the system may still be opaque if the rules are not general or have numerous exceptions. One example is Manambu (described further below), where gender assignment sometimes carries the notion of large size, so that larger animals are masculine and smaller animals feminine. However, insects are masculine despite their small size.<sup>10</sup>

It is therefore possible to further split the systems with semantic assignment into two: transparent semantic vs. semantic + opaque (Figure 5), where opacity signals the inability of the researcher to find any clear semantic or formal criteria for gender assignment. It is possible that a language may have semantic + formal + opaque assignment, but no such system was clearly identified in the sample.

---

<sup>10</sup>It is of course possible to imagine various explanations why insects are not feminine, e.g., perhaps are they are not regarded as animals. However, this only further illustrates the reason for not regarding Manambu gender assignment as transparent. Although there certainly is a general pattern of size distinctions for gender assignment in Manambu, it is merely a pattern and not a rule.

Table 4: Systems of gender assignment in the sample

Gender assignment	No. of lgs.	%	Languages
Semantic	16	80%	Abau
			Ama
			Au
			Burmeso
			Manambu
			Maybrat
			Mende
			Mian
			Motuna
			Oksapmin
			Rotokas
			Skou
			Taiap
			Teop
			Walman
			Warapu
Semantic and formal	4	20%	Bukiyip
			Kuot
			Nalca
			Yimas
Total:	20	100%	

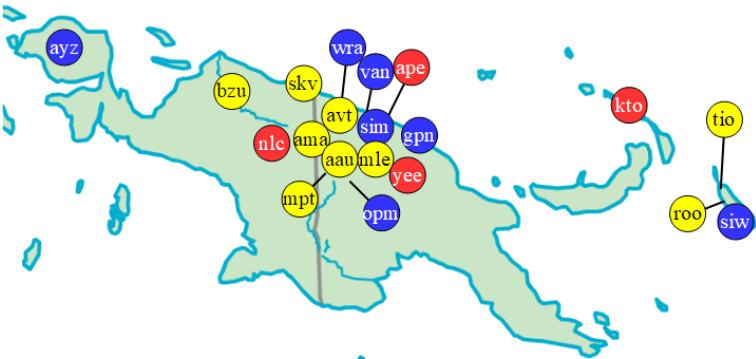


Figure 4: Systems of gender assignment. Colors indicate transparent semantic (blue), semantic and formal (red), and semantic and opaque (yellow).

Table 5: Types of semantic assignment of gender assignment in the sample

Gender assignment	No. of lgs.	%	Languages
Transparent semantic	8	40%	Au Maybrat Mende Motuna Oksapmin Taiap Walman Warapu
Semantic and formal	4	20%	Bukiyip Kuot Nalca Yimas
Semantic and opaque	8	40	Abau Ama Burmeseo Manambu Mian Rotokas Skou Teop
Total:	20	100%	

This thus gives rise to three types of gender assignment: transparent semantic, semantic and formal, and semantic and opaque (Table 5).<sup>11</sup>

Since all languages have some form of semantic assignment, the most basic system is necessarily one where all nouns are assigned their genders based on few and clear semantic criteria. Mende has already been mentioned above and is exemplified in (1) in §3.1. However, semantic systems can be more complex while still retaining transparent semantic criteria, e.g., via a larger number of

<sup>11</sup>It is not explicitly stated, but Au (Scorza 1985) appears to have a simple semantic system where nouns denoting human males are masculine, human females are feminine, and the rest are neuter. However, this is complicated somewhat by masculine and neuter agreement being homophonous in the singular.



Table 6: Gender indexation forms in Motuna (adapted from Onishi 1994: 70)

	Demonstrative	Article	Adjective/ classifier/ kinship term endings	Possessor/ local NP endings	Verbal endings
Masculine	<i>ong</i>	<i>hoo/shoo</i>	<i>-ng</i>	<i>-ng</i>	<i>-ng</i>
Feminine	<i>ana</i>	<i>tii</i>	<i>-na</i>	<i>-na</i>	<i>-na</i>
Diminutive	<i>oi</i>	<i>tii</i>	<i>-ni</i>	<i>-ni</i>	<i>-ni</i>
Local	<i>owo</i>	<i>ti</i>	–	<i>-no</i>	<i>-no</i>
Manner	–	<i>tiwo</i>	–	–	<i>-nowo</i>
Dual-paucal	<i>oi</i>	<i>tii</i>	–	<i>-ni</i>	<i>-(n)i</i>

gender distinctions. One example is Motuna (South Bougainville), which has six genders: masculine, feminine, diminutive, local, manner, and dual-paucal (Onishi 1994: 68–69). The forms of gender indexation in Motuna are shown in Table 6.

In Motuna, animate referents are assigned gender based on their natural gender; this also includes nouns associated with mythical characters such as *raa* ‘the sun’ and *hingjoo* ‘the moon’, which are assigned the gender of their character (Onishi 1994: 70). Animals are most commonly masculine, but can be assigned the feminine gender if emphasizing that the referent is a female. On the other hand, the majority of inanimate nouns are masculine, but can be treated as diminutive when emphasis is placed on their small size. This includes nouns which signify smallish things, e.g., *irihwa* ‘finger’ or *kaa* ‘young tree’ (Onishi 1994: 71). Nouns with spatial or temporal meaning are inherently local gender. The manner gender contains only two nouns. Finally, the dual-paucal gender can be used also when the speaker does not want to specify the gender of a sentential topic (Onishi 1994: 71).

In contrast to the transparent semantic criteria in Mende and Motuna mentioned above, many languages have much more complex systems. If gender assignment is neither semantically transparent nor apparently formal, it is classified as being opaque, with Rotokas having already been mentioned at the beginning of this section. Another example of such a language is Manambu (Ndu), which exhibits the fairly common feature of gender assignment based on size and shape (see §5.1). Manambu has two genders, masculine and feminine, and in general gender assignment appears to follow semantic criteria. However, these are far from transparent:

replaced  
“

1. Humans are assigned gender based on their sex, except nouns denoting small children, which can be assigned gender based on size (Aikhenvald 2008: 116–117).
2. Higher animates are assigned gender based on their size and natural gender: larger animals are masculine, whereas smaller animals are feminine, except when the sex of the referent is known. Furthermore, nouns denoting young animals are feminine (Aikhenvald 2008: 117).
3. Lower animates such as insects are masculine. However, if the lower animate has a certain shape, it is assigned gender based on shape; thus, *gwa:s* ‘turtle’ is feminine, since it is round, while *mu* ‘crocodile’ is masculine since it is long (Aikhenvald 2008: 117).
4. Inanimates are assigned gender based on their size and shape: long and/or large objects are masculine, whereas small and/or round objects are feminine. Thus, *vø:y* ‘spear’ is masculine, since it is large, but it is feminine if referring to small spears or shotguns (Aikhenvald 2008: 117).
5. Natural phenomena are assigned gender based on whether they are complete or not: if they are uncomplete or if completeness is not emphasized, they are feminine; otherwise, they are masculine (Aikhenvald 2008: 118). Thus, *ga:n* ‘night’ is feminine, unless it implies complete darkness, as is *gəl* ‘cloud’ if there are only a few but masculine if they cover the whole sky. Other natural phenomena are assigned gender based on their shape: e.g., ‘rainbow’ is masculine since it is long, whereas ‘sun’ is feminine since it is round; unless it is really hot, in which case it becomes masculine to reflect its intensity (Aikhenvald 2008: 119).
6. Mass nouns and nouns covering ‘extent’ follow complex patterns. In general, they are assigned gender based on extremity, so that smaller quantities are feminine, whereas larger quantities are masculine (Aikhenvald 2008: 119–120). However, nouns denoting manner, language or voice, or time span are feminine; except for *nabi* ‘year’, which is masculine because it is very long (Aikhenvald 2008: 119).

There are in fact further assignment rules, but the important point is that rules of gender assignment are not semantically transparent. It is especially important to note that it is difficult to ascertain whether there are any rules or merely patterns. That is not to belittle the observations or to claim that the researcher, in

this case Aikhenvald, has done anything wrong. Instead it illustrates the difference to transparent semantic systems, where all gender assignment rules are easily identifiable and apply to all nouns, whereas in opaque systems there are certainly some patterns that can be identified, but exceptions abound.

While it is easy to become amused by the seemingly arbitrary gender assignment rules, one important thing should be noted. In a language such as Manambu, gender has a very important pragmatic function, since it is available as a tool for the speaker to use when emphasizing certain features, not least in jokes:

As a joke, a man can be referred to with feminine gender, and a woman with masculine gender, depending on their ‘shape’ and ‘size’. A smallish fat woman-like man can be treated as feminine, e.g. *numa du* (big.FSG man) ‘fat round man’. And a largish woman can be ironically referred to with a masculine gender form, e.g. *kə-də numa-də ta:kw* (DEM.PROX-M.SG big-M.SG woman) ‘this (unusually) big woman’. (Aikhenvald 2008: 121)

The final category consists of the four languages with both semantic and formal assignment: Nalca is skewed towards semantic assignment, Kuot applies semantic and formal assignment roughly equally, and Bukiyip and Yimas favor formal assignment. For example, among the five genders in Nalca (see §3.1 above), only the neuter is formal, but very much so since it contains only (but all) nouns of the phonological structure (C)V (Wälchli 2018). In comparison, only three of the 18 genders in Bukiyip (see Table 10) are semantic (masculine, feminine, and mixed or unspecified), whereas all others are morphological (Conrad & Wogiga 1991: 8). The same is true for Yimas, where three genders are semantic, while the others are based on phonological criteria (Foley 1991: 119).

### 3.4 Number of gender-indexing targets

Following Di Garbo (2014: 66), the number of gender-indexing targets is given the value of one, two, three, or four or more. The results are shown in Table 7 and Figure 5, while each type of indexing target is shown in Table 8. The identification and counting of gender-indexing targets was based on the general guidelines used by Di Garbo (2014: 66), where the following general categories were used to identify targets: pronouns, adjectives, demonstratives, verbs, numerals, copulas, complementizers, and adpositions. However, no detailed analysis has been made of different subtypes of these groupings, so the results should be understood only as showing general patterns.

Table 7: Number of gender-indexing targets in the languages in the sample

Number of gender-indexing targets	No. of lgs.	%	Languages
One	4	20%	Ama Mende Nalca Oksapmin
Two	4	20%	Warapu Burmeso Skou Teop
Three	2	10%	Au Taiap
Four or more	10	50%	Abau Bukiyip Kuot Maybrat Manambu Mian Motuna Rotokas Walman Yimas
Total:	20	100%	



Figure 5: Number of gender-indexing targets. Colors indicate: one (blue), two (green), three (yellow), and four or more (red).

Table 8: Distribution of gender-indexing targets in the languages of the sample

Language	Pronouns	Verbs	Demonstratives	Adjectives	Numerals	Prepositions	Case hosts
Abau	x		x		x		x
Ama		x					
Au	x	x		x			
Bukiyip	x	x	x	x			
Burmeso		x		x			
Kuot	x	x	x	x		x	
Manambu	x	x	x	x			
Maybrat	x	x	x	x			
Mende	x						
Mian	x	x	x	x			
Motuna	x	x	x	x	x		
Nalca							x
Oksapmin	x						
Rotokas	x	x	x	x			
Skou	x	x					
Taiap	x	x	x				
Teop				x	x		
Walman	x	x	x	x	x		
Warapu	x	x					
Yimas	x	x	x	x	x		

As shown in Table 7 and 8, more than half of the languages in the sample have more than four gender-indexing targets.<sup>12</sup> There are also some general patterns to be found in Table 8:<sup>13</sup>

<sup>12</sup>In Burmeso, adjectives are targets in the first gender system whereas verbs are targets in the second system.

<sup>13</sup>'Pronoun' here denotes a word with general pronominal uses (i.e., as constituting an individual noun phrase), whether it belongs to the language-specific category of pronouns or demonstratives. In comparison, 'demonstrative' only refers to attributive forms. Kuot has no independent third person personal pronouns (Eva Lindström, p.c.). However, demonstratives are used with pronominal functions (see also footnote 16 below). 'True pronouns' in Yimas exist only in the first and second person without gender (Foley 1991: 111). The third person is instead expressed with a set of deictics, which show gender and are most commonly used as free pronouns in narrative discourse (Foley 1991: 113). Therefore these forms are considered pronouns for comparative purposes.

- If a language has four gender-indexing targets, they always include pronouns and demonstratives, and almost all such languages include verbs and adjectives, with Abau (Sepik) being the exception.
- If a language has three gender-indexing targets, they include verbs and pronouns.
- If a language has two gender-indexing targets, they mostly include verbs and to a lesser extent pronouns.
- If a language has only one gender-indexing target, the target could be anything (e.g., verbs, pronouns, or even case marker hosts).

Based on the likelihood of a gender-indexing target appearing in a language, it is possible to arrange the distributional tendencies into tentative hierarchies, where the leftmost target is the most typical target while the rightmost target is the least common one. If one target is present in a language, every target to the left is present as well. That is, if a language has only one target, it is likely to be the leftmost one, whereas if a language has five it should include every part of the hierarchy. There are three tendencies:

- pronouns > verbs > demonstratives > adjectives > numerals (holds for 14 out of 20 languages)
- verbs > adjectives > pronouns (3/20)
- other (3/20)

It is also interesting to note that among the ten languages with four or more indexing targets, all except Abau follow the first hierarchy. There is therefore an additional pattern, whereby a gender system with many indexing targets is expected to follow the first hierarchy. In comparison, four of the six languages of the other two categories have two gender-indexing targets or less, with Au having three targets and Abau four.

The languages not describable in terms of the first and second hierarchies are all very different and require some explanation. One example is Nalca (TNG, Mek), which only shows gender on markers functioning as case marker hosts following the NP. These carry the meaning of gender, case, and demonstrative, of which at least the first two mostly occur together. Some of the most common forms are shown in Table 9. Examples were given in (3) in §3.1 above, the first of which is repeated in (5).

Table 9: Some of the most of most frequent forms of case marker hosts words in Nalca.

Case	masc. <i>be-</i>	fem. <i>ge-</i>	neuter <i>ne-</i>	default noun <i>e-</i>	non-noun <i>a-</i>
Topic	<i>bera</i>	<i>gera</i>	<i>nera</i>	<i>era</i>	<i>ara</i>
Topic dem.	<i>benera</i>	<i>genera</i>	<i>nenera</i>	<i>enera</i>	<i>anara/anera</i>
Absolutive	<i>bek</i>	<i>gek</i>	<i>nek</i>	<i>ek</i>	<i>ak</i>
Abs. dem.	<i>benyek</i>	<i>genyek</i>	<i>nenyek</i>	<i>enyek</i>	<i>anyek</i>
Gen./ergative	<i>bedya(')</i>	<i>gedya(')</i>	<i>nedya(')</i>	<i>edya(')</i>	<i>adya(')</i>
Gen./erg. dem.	<i>benedya</i>	<i>genedya</i>	<i>nenedya</i>	<i>enedya</i>	<i>anadya</i>
Comitative	<i>beb</i>	<i>geb</i>	<i>neb</i>	<i>eb</i>	<i>ab</i>
Com. dem.	<i>benyeb</i>	<i>genyeb</i>	<i>nenyeb</i>	<i>enyeb</i>	<i>anyeb</i>
Equative	<i>beneso(')</i>	<i>geneso(')</i>	<i>neneso(')</i>	<i>eneso(')</i>	<i>anaso(')</i>
Benefactive	<i>bemba</i>	<i>gemba</i>	<i>nemba</i>	<i>emba</i>	<i>amba</i>

- (5) Nalca (TNG, Mek) (own example; repeated from 3a)

*alja si ne-ra Zakheus be-k u-lum-ok*  
 3SG.GEN name N-TOP Z. M-ABS be-IPFV-PST.3SG  
 ‘a man called by name Zacchaeus’ (Lk 19:2)  
 lit. ‘his name was Zacchaeus’

Another interesting example is Teop (Austronesian, Oceanic). In Teop, gender is visible on a set of articles preceding nouns, adjectives, and numerals. Two examples of markers preceding adjectives and numerals, respectively, are shown in (6).

- (6) Teop (Austronesian, Oceanic) (Mosel & Spriggs 2000: 330, 328)

- a. *a inu a beera*  
 ART.I.SG house ART.I.SG big  
 ‘the big house’
- b. *o buaku o hoi*  
 ART.II.SG two ART.II.SG basket  
 ‘the two baskets’

However, since these articles do not carry any other functional load, they do not satisfy the criterion that an indexing target must express something other

than gender and number. Instead, Teop is analyzed as having two targets, viz., adjectives and numerals, which form a unit with the preceding article. On the other hand, the articles preceding nouns are analyzed as overt gender marking (see §3.5).

### 3.5 Occurrence of gender marking on nouns

The final criterion concerns the occurrence of gender marking on nouns (see Table 10 and Figure 6), following Di Garbo (2014: 69). Gender marking on nouns is of course not considered indexation, but it is a common feature e.g. in African languages and most certainly a characteristic trait of many gender systems.

Most languages of the sample (17 of 20) do not have overt gender marking, with Bukiyip, Teop, and Yimas being the only exceptions. In both Bukiyip and Yimas, gender is shown on nouns via suffixes; the Bukiyip noun suffixes are given in Table 11. Both languages are unusual in the sample by their having many noun classes (18 in Bukiyip, around a dozen in Yimas), many gender-indexing targets (both four or more), and semantic-formal assignment. In fact, these features are probably tightly interconnected with the overtness of gender. The combination of many genders and morphological gender assignment appears more common when noun classes are overtly distinct.

On the other hand, Teop (Austronesian, Oceanic) has a very different kind of marking. As mentioned above, Teop has a set of articles that obligatorily precede nouns, adjectives, and numerals. Thus, the latter two are indexation, while the articles preceding nouns are considered overt marking. The forms of the markers are shown in Table 12.

Note that Teop has two genders, one of which is divided into two subgenders. The reason for them not being separate gender is that the distinction is kept only on the articles preceding nouns, and never on the articles preceding adjectives and numerals. Thus, since overt gender marking cannot constitute gender as it is not indexation, Teop only has two genders.

This is very similar to the related Austronesian language Siar (not in the sample), which also has articles preceding nouns (Frowein 2011). However, the Siar articles are not used in other contexts, so the absence of indexation renders Siar genderless. Nevertheless, a pronoun can be placed before e.g., an adjective, which is similar to the use of the Teop article. However, pronouns in Siar do not show any gender distinctions. The difference between Teop and Siar in this regard is shown in (7) and (8), respectively.



Table 10: Occurrence of gender marking on nouns in the sample

Gender marking on nouns	No. of lgs.	%	Languages
Yes	3	15%	Bukiyip Teop Yimas
No	17	85%	Abau Ama Au Burmeso Kuot Manambu Maybrat Mende Mian Motuna Nalca Oksapmin Rotokas Skou Taiap Walman Warapu
Total:	20	100%	

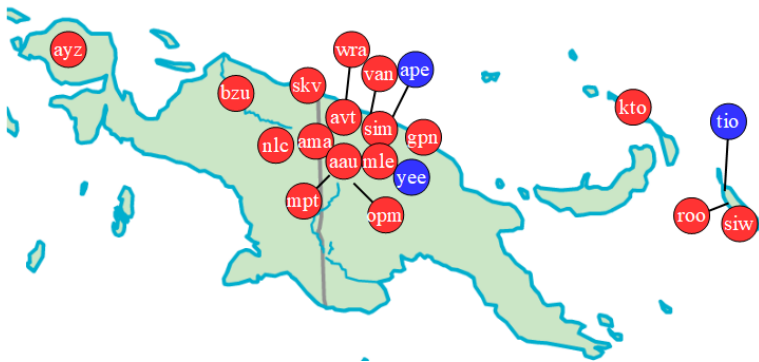


Figure 6: Occurrence of gender marking on nouns. Colors indicate: yes (blue), and no (red).

Table 11: Bukiyip noun classes and noun class suffixes (adapted from Conrad &amp; Wogiga 1991: 10)

Class	Glossing	Example		Noun suffix	
		singular	plural	singular	plural
1	betel nut	<i>búb</i>	<i>búbús</i>	<i>-b/n</i>	<i>-bús</i>
2	village	<i>wabél</i>	<i>walúb</i>	<i>-bél</i>	<i>-lúb</i>
3	feces	<i>dewag</i>	<i>dewas</i>	<i>-g/-gú</i>	<i>-s/-as</i>
4	woman	<i>élmatok</i>	<i>élmagou</i>	<i>-k</i>	<i>-ou/-eb</i>
5	banana	<i>apam</i>	<i>apas</i>	<i>-m/-bal</i>	<i>-s/-ipi/-bal</i>
6	moon	<i>aun</i>	<i>aub</i>	<i>-n/-nú</i>	<i>-b</i>
7	man	<i>élman</i>	<i>élmom</i>	<i>-n/-nú</i>	<i>-m</i>
8	child	<i>batawiny</i>	<i>batawich</i>	<i>-ny/-l</i>	<i>-ch/-has</i>
9	leaf	<i>chuwup</i>	<i>chuwus</i>	<i>-p</i>	<i>-s</i>
10	mosquito	<i>aul</i>	<i>auguh</i>	<i>-l/-ny</i>	<i>-guh</i>
11	dog	<i>nobat</i>	<i>nobagw</i>	<i>-t/-tú</i>	
12	sago leaves	<i>lohuhw</i>	<i>lohulúh</i>	<i>-hw</i>	
13	road	<i>yah</i>	<i>yeh/yegwih</i>	<i>-V<sub>1</sub>h</i>	<i>-V<sub>2</sub>h</i>
14	box		<i>kes</i>	<i>-s</i>	<i>-s</i>
15	small pig		<i>buligún</i>	<i>-gún</i>	<i>-gún</i>
16	garden		<i>yawihas</i>	<i>-has</i>	<i>-has</i>
17	personal names			<i>-</i>	<i>-</i>
18	place names			<i>-</i>	<i>-gún</i>

Table 12: Gender marking in Teop on articles preceding nouns (Mosel &amp; Spriggs 2000: 322)

	head (SG)	head (PL)	target (SG)	target (PL)
Gender I-E	<i>e</i>	<i>o</i>	<i>a</i>	<i>o</i>
Gender I-A	<i>a</i>	<i>o</i>	<i>a</i>	<i>o</i>
Gender II	<i>o</i>	<i>a</i>	<i>o</i>	<i>a</i>

- (7) Teop (Austronesian, Oceanic) (Mosel & Spriggs 2000: 326)

*a inu a ruta*

ART.I house ART.I small

‘the small house / the house is small’

- (8) Siar (Austronesian, Oceanic) (Frowein 2011: 206)

*Ēp rumai i mètək.*

ART.CO1 house 3SG new

‘The house is new.’

Finally, some languages have overt marking in some cases or at least something resembling it. One example is Kuot (isolate), where some nouns belong to various declension classes (as defined by noun endings), which in turn belong to a certain gender (Lindström 2002: 176). Another example is Rotokas (North Bougainville), which has noun suffixes expressing both number and gender (Robinson 2011: 41). However, these are not always present: in (9a), *aveke* ‘stone’ has a feminine singular suffix, but in (9b) it remains unmarked.

- (9) Rotokas (Robinson 2011: 42)

a. *riako-va aveke-va peka-e-vo uva rakoru*

woman-SG.F stone-SG.F turn.over-3SG.F-IPST and snake

*keke-e-vo uva kea-o-e oisio uo-va*

look.at-3SG.F-IPST and mistake.for-3SG.F-IPST as eel-SG.F

‘The woman turned over to the stone and saw a snake but mistook it for an eel.’

b. *kaveakapie-vira aveke tovo-i-vo uva kove-o-e*

insecure-ADV stone place-3PL-IPST and fall-3SG.F-IPST

‘They placed the stone insecurely and it fell down.’

Since gender marking on nouns is not always present, Rotokas cannot be said to have obligatory overt marking.

## 4 Typological comparison

This section compares the results of this study with previous research on Africa and the world as a whole. The data on Africa is from Di Garbo (2014), which used the same five criteria of this study to investigate a variety sample of 100

languages. The data on the world as a whole is based on the three *WALS* chapters on gender by Corbett (2013a,b,c). These three *WALS* chapters correspond to the first three classification criteria of this study. Unfortunately, the remaining two have no corresponding *WALS* data, rendering the final two criteria comparable only for New Guinea and Africa.

Some care had to be taken when comparing the results, since the samples are of different types. Whereas this study employs a variety sample, Corbett uses a proportional sample (of 257 languages) (see §2). Di Garbo also uses a variety sample (of 100 languages) although with some differences, most importantly the inclusion of 16 non-gendered languages as well as being intentionally genealogically skewed. To make the data comparable, languages without gender have been omitted from Corbett's and Di Garbo's samples in this section, leaving 112 languages for Corbett and 84 for Di Garbo.

*Classification criterion 1: Sex-based and non-sex-based gender systems* (§3.1). In the sample of this study, sex-based systems are by far more common, with only Teop (Austronesian, Oceanic) having a non sex-based system. In comparison, in Di Garbo's (2014: 63) sample, 48 languages (57%) had sex-based gender systems and 36 languages (43%) non-sex-based gender systems. In Corbett's (2013c) sample, 84 languages (75%) have sex-based systems and 28 (25%) non-sex-based. A comparison of the percentage distributions is shown in Figure 7.

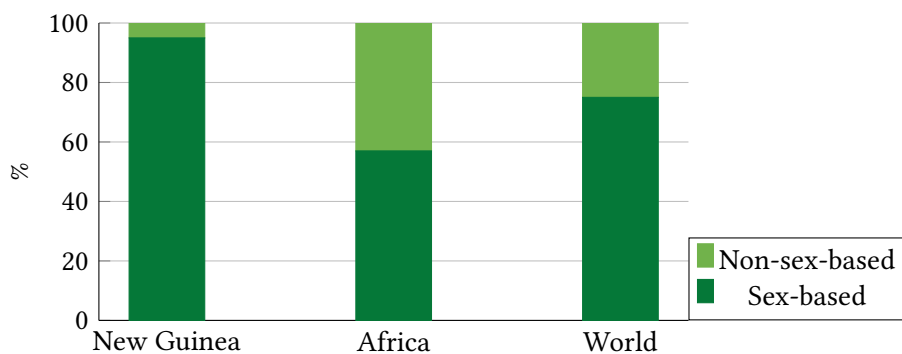


Figure 7: Sex-based and non-sex-based systems in New Guinea, Africa and the world

Sex-based systems are more common in all samples, although even more so in the sample from New Guinea. According to Corbett's (2013b) data, non-sex-gender systems are actually uncommon in most regions, being found primarily in the Niger-Congo languages of Africa, which account for the vast majority of non-sex-based systems in the sample. More specifically for Africa, in most cases

only one system occurs in an entire family: this is true e.g., for the Bantu, Mel, and North-Central Atlantic families, which together account for 33 of the 36 non-sex-based gender systems in Di Garbo's sample. It is therefore not surprising that the non-sex-gender systems are relatively common, since 31% of the gendered languages (26/84) in Di Garbo's sample are Bantu languages.

An interesting discussion about the differences between sex-based and non-sex-based systems is presented by Luraghi (2011), who argues that they have different diachronic origins. Non-sex-based systems originate from the grammaticalization of classifiers, whereas sex-based systems originate from agreement with groups of nouns that show different morphosyntactic behavior. Since sex-based systems are more common, it is thus not surprising that they are the primary ones in New Guinea. It is likely not a coincidence that the only non-sex-based gender system of the sample is found in an Austronesian language, a family remarkably devoid of gender but abounding with classifiers.

*Classification criterion 2: Number of genders* (§3.2). In the sample of this study, eleven languages (55%) have only two genders, three languages (15%) three genders, one language (Mian; TNG, Ok-Oksapmin) (5%) four genders, and the final five languages (25%) five genders or more. In Di Garbo's (2014: 65) sample, 42 languages (50%) have only two genders, seven languages (8%) three genders, one (Jul'hoan; Kx'a) (1%) four genders, and the final 34 languages (40%) five genders or more. In Corbett's (2013a) sample, 50 languages (45%) have only two genders, 26 languages (23%) three genders, 12 languages (11%) four genders, and the final 24 (21%) five genders or more. A comparison between the percentage distributions is shown in Figure 8.

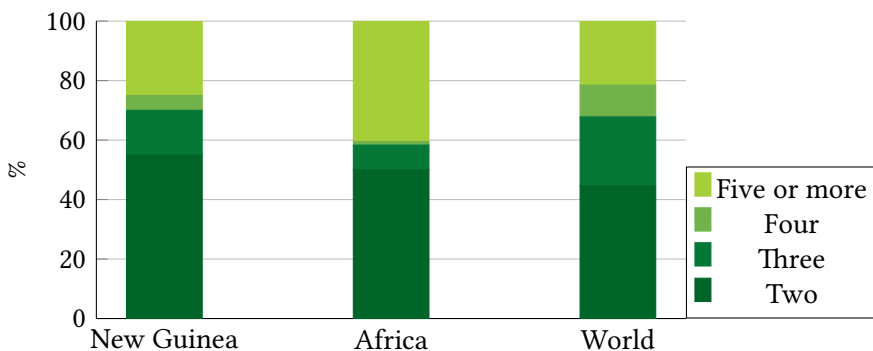


Figure 8: Number of genders in New Guinea, Africa and the world

The distributions in all three samples are similar to a large extent, with two-gender systems being present in around half of the languages. In Africa, large

systems are much more common than in New Guinea or the world as a whole. However, this may once again be because of the sample. As mentioned before, 31% of the languages present in Di Garbo's (2014) sample are Bantu languages, all of which have very large gender systems. In the sample of this study, however, the rather large Torricelli and Lower Sepik-Ramu families, which according to Foley (2000: 372) have large systems, are represented only by Bukiyip and Yimas respectively (i.e., 10% of the sample). It is thus very probable that the similarities between the distribution numbers of genders in New Guinea and Africa actually are greater than indicated here.

*Classification criterion 3: Gender assignment (§3.3).* This criterion is less straightforward to compare, since this study uses three values (transparent semantic, semantic and formal, and opaque), whereas Di Garbo (2014) and Corbett (2013c) use only two (semantic, and semantic and formal). For the purpose of this comparison, the languages of the purely semantic and semantic + opaque groups are added somewhat tentatively into a semantic group. While this may appear misleading, it is important to note that the researchers investigating these languages considered them as having semantic gender assignment and no traces of formal assignment rules have been identified by the present author. Indeed, both languages exemplified in Corbett (2013c), Bininj Gun-Wok (Gunwinyic; northern Australia) and Russian, would be considered opaque using the values of this study.

In the sample of this study, 16 languages (80%) exhibit semantic gender assignment, whereas only four languages (20%) show semantic and formal assignment. In comparison, in Di Garbo's (2014: 67) sample, six languages (7%) have semantic assignment, 76 languages (90%) semantic and formal assignment, while the remaining two languages (2%) have unknown assignment (disregarded in Figure 9). In Corbett's (2013c) sample, 53 languages (47%) exhibit semantic assignment, and 59 languages (53%) semantic and formal assignment. A comparison between the percentage distributions is shown in Figure 9.

As can be clearly seen in Figure 9, in New Guinea, semantic assignment is by far more common, while it is by far the most uncommon form of gender assignment in Africa, including of course the Bantu languages. In the world as a whole, the ratio is more or less equal. Thus, New Guinea and Africa both represent two extremes while the world as a whole is more average. However, according to Corbett (2013c), semantic and formal assignment is mostly found in the Indo-European, Afro-Asiatic, and Niger-Congo families, which together represent a large amount of the languages of the world.

It is not surprising that semantic and formal assignment appears more often in Di Garbo's and Corbett's samples than in New Guinea, since no family is rep-

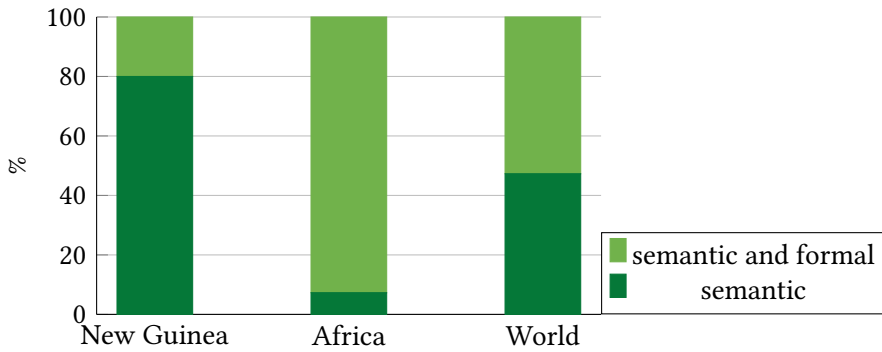


Figure 9: Gender assignment in New Guinea, Africa and the world

resented with more than three members in this study. Bukiyip (Torricelli, Arapesh) and Yimas (Lower Sepik-Ramu, Lower Sepik) both belong to rather large families, so it is possible that a proportional sample would show that semantic and formal assignment indeed is more common than it appears here. Nevertheless, it is interesting that it occurs in few families, both in New Guinea and the world, which Corbett (2013c) relates to these systems necessarily being older. As argued by Luraghi (2011), this implies that most gender systems of Africa are old. Exclusive semantic assignment is however found in both older and younger systems, and thus it cannot be claimed that the predominance of semantic assignment indicates that those systems are young. Interestingly, semantic and formal assignment is found in Nalca (TNG, Mek), which has a very young gender system (Wälchli 2018).

*Classification criterion 4: Number of gender-indexing targets (§3.4).* In the sample of this study, four languages (20%) have only one gender-indexing target, another four languages (20) two targets, two languages (10%) three targets, and the final ten languages (50%) four or more targets. In Di Garbo's (2014: 68) sample, five languages (6%) have only one gender-indexing target, 16 languages (19%) two targets, 28 languages (33%) three targets, and finally 33 languages (39%) four targets or more. No data was available for the remaining two languages. A comparison of the percentual distributions is shown in Figure 10.

Four or more gender-indexing targets is the most common number in both samples, accounting for slightly less than half of all languages. Furthermore, systems of only two targets account for around a fifth of the languages in both samples. As for the two remaining values, the relationships are the opposite: systems of three targets are common in Africa but rare in New Guinea, whereas one-target systems occur in a fifth of the New Guinean languages but only 6% of

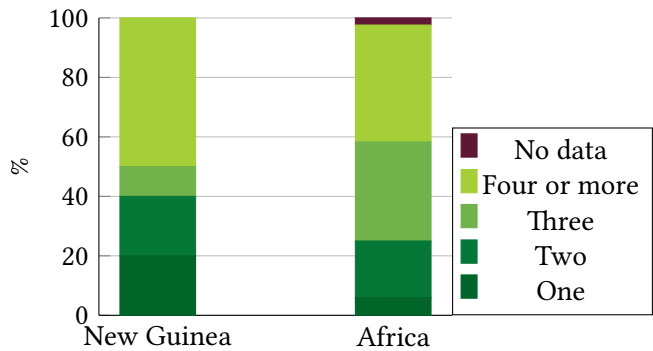


Figure 10: Number of gender-indexing targets in New Guinea vs. Africa

the African languages. However, once again it is probable that these differences are largely due to larger families with more mature gender systems being better represented in Di Garbo’s (2014) sample, while languages from smaller families with possibly less mature gender systems constitute a large part of the sample of this study.

*Classification criterion 5: Occurrence of gender marking on nouns (§3.5).* In the sample of this study, three languages (15%) have overt gender marking, whereas the remaining 17 (85%) do not. In Di Garbo’s sample, 69 languages (82%) have overt gender marking and 15 (18%) do not. A comparison between the percentage distributions is shown in Figure 11.

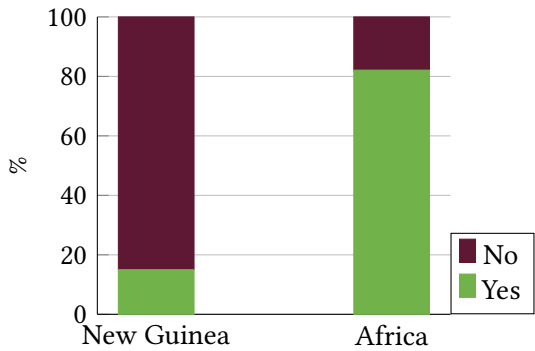


Figure 11: Occurrence of gender marking on nouns in New Guinea vs. Africa

As the figure shows, there is a major disparity between the presence of gender marking on nouns in New Guinea and Africa. In New Guinea, overt gender



marking is rare and occurs in only three languages in the sample, whereas in Africa it occurs in the vast majority of languages.

There is an interesting correlation between this distribution and the one of gender-assignment shown in Figure 4. Thus, semantic assignment without gender marking on nouns is the norm in New Guinea, whereas semantic and formal assignment with gender marking on nouns is the norm in Africa. This correlation is hardly coincidental. A gender system with assignment based on formal criteria benefits greatly from overt gender. In an exclusively semantic system however, obligatory overt gender has no function in gender assignment.

To summarize, it can be confidently stated that the gender systems of New Guinea and Africa are very different. Much of this depends on the hegemony of Bantu languages in Africa (as represented by Di Garbo's sample), which makes the distribution of values much less diverse than in the sample of this study. Nevertheless, the most important differences are (1) the prevalence of semantic and formal assignment and overt gender in Africa, while the exact opposite is true in New Guinea, and (2) as the observation that non-sex-based genders are much more common in Africa. This clearly shows that the two regions have gender systems of very different types. Reasons for this definitely include sample size and technique, but it also suggests that the gender systems of New Guinea may have different diachronic origins.

As for New Guinea in relation to the world as a whole, the above data and figures show that the distribution of values of the three classification criteria is rather similar in New Guinea and the world. In fact, most of the smaller differences can probably be accounted for by sample size. Nevertheless, the main conclusion is that the languages of New Guinea seem to be remarkably representative of the languages of the world, but another study with a proportional sample from New Guinea would elucidate this further.

## 5 Special characteristics

In this section, four characteristics of the gender systems of New Guinea are highlighted, two of which reflect characteristics mentioned by Foley (2000), viz., gender assignment based on size and shape, and the occurrence of two separate gender systems. The other two, viz., no gender distinctions in pronouns and gender marking on verbs, pertain to two typologically uncommon characteristics. Although these do not occur in all languages of the sample, they are found in geographically and genealogically distant languages and are all characteristic of the region.

## 5.1 Size and shape

Four languages in the sample (20%) share the property of having size and shape as important criteria for gender assignment. While gender assignment in many languages may carry some form of size- or shape-based rules, the rules discussed here all share the feature that nouns denoting tall, long, or thin objects are considered masculine, whereas nouns denoting short, thick, or round objects are feminine. In addition, they are all core assignment criteria. The languages in the sample exhibiting this feature are: Abau (Sepik), Manambu (Ndu), Skou (Sko), and Taiap (isolate).<sup>14</sup> Their rules based on shape and size are shown in Table 13.

Table 13: Gender assignment rules based on size and shape in the sample

Language	Masculine	Feminine
Abau	- large - three-dimensional - long and extended	- small - two-dimensional (i.e., very thin) - round with little height
Manambu	- large - long	- small - round
Skou	- large - long, thin	- small - round, squat
Taiap	- large - long, high, thin	- small - round, stocky

In these four languages, size and shape are important criteria for gender assignment. One example mentioned in §3.1 above is Abau, which has two genders: masculine and feminine. Humans, along with spirits and domesticated animals, are assigned gender based on their sex, whereas abstract entities are feminine (Lock 2011: 47). However, animals and concrete inanimate objects are assigned their gender based on shape and size. Large, three-dimensional, and/or long and extended objects are masculine, while small, two-dimensional (i.e., very thin), and/or round objects with little height are feminine (Lock 2011: 47). Thus, *su* ‘coconut’ (three-dimensional), *now* ‘tree’ (long), and *hu* ‘water’ (liquid) are masculine, while *iha* ‘hand’ (flat) and *hne* ‘bird’s nest’ (round with little height) are feminine (Lock 2011: 48–50).

<sup>14</sup>‘Non-feminine’ in Skou.

It is important to distinguish systems such as the ones above from diminutives. In some languages, diminutives constitute separate genders, such as in Motuna (South Bougainville) (Onishi 1994: 68–69). However, the four languages above show the peculiar characteristics that (1) size and shape function as assignment criteria for the masculine and feminine genders, and (2) they constitute opposing criteria, and (3) they show the same pattern of large/long vs. small/round.

In the sample, size and shape constitute important gender assignment criteria in only these four languages, but similar systems are present in other languages. Rotokas exhibits some similarities with these gender assignment rules in two ways. Firstly, one class of nouns belonging to the masculine gender consists of inanimate objects associated with male culture, but also includes long or thin objects. However, no comparable feminine gender assignment rule has been found. Furthermore, this appears to be only a peripheral gender assignment rule. Secondly, Rotokas has a set of classifiers based on shape and size, classifying nouns based on their being round, narrow, or long. While this is not related to any masculine-feminine opposition, it nonetheless bears some resemblance to these systems.

Another interesting example is Mian (TNG, Ok). Mian has four genders, viz., male, female, neuter 1, and neuter 2, none of which has gender assignment rules resembling those of size and shape (Fedden 2011: 171–176). However, around 50 verbs require the use of a classificatory prefix, which has two functions: firstly, it encodes the direct object of transitive verbs and the subject of intransitive verbs, and secondly it classifies it according to characteristics of the referent, viz., sex, shape, and function (Fedden 2011: 185). This classification system, which is separate from the gender system, includes classes for e.g., long or flat objects, and in some cases overlaps with the gender system (e.g., some neuter 1 nouns are included in the masculine class). A table illustrating the overlap between the two systems is shown in Table 14.

Assigning genders based on shape and size is not very common in the languages of the world (Aikhenvald 2000: chap. 11). Outside of New Guinea, it occurs e.g., in some Afroasiatic languages, such as Oromo and Amharic, Central Khoisan, and Cantabrian Spanish (Aikhenvald 2000: 277; Heine 1982: 191). However, size as an assignment criterion is widespread in Africa, where it e.g., occurs in diminutive and augmentative genders as reported by Di Garbo (2014). An example is in Tonga, where ‘boy’ (noun class 1) can shift to the diminutive noun class 12 to highlight smallness:

Table 14: Overlap between the gender and verb prefix classes of Mian (adapted from Fedden & Corbett 2017: 34). Cells with examples show the attested combinations.

	Masculine	Feminine	Neuter 1	Neuter 2
M-classifier	man, boy, boar	–	sleeping bag, plate, mosquito net	–
F-classifier	–	woman, girl, sow	–	house, steel axe, money
Long	–	–	tobacco, eating implement, bush knife	–
Bundle	–	–	string bag, plastic bag	–
Covering	–	–	blanket, band aid	–
Residue	–	tortoise, scorpion	cassowary egg, plane, hat	–

(10) Tonga (Bantu) (Di Garbo 2014: 147; from Carter 2002: 21)

- a. *mu-sankwa*  
CL1-boy  
'boy'
- b. *tu-sankwa*  
CL12-boy  
'small boy'

As for New Guinea, its prevalence specifically in the Sepik area has led Aikhenvald (2008: 113) to suggest that gender assignment based on size and shape may actually be an areal feature of the Sepik area. Indeed, all four languages in this sample found to have such systems are spoken in or near the Sepik area: Abau (Sepik) and Manambu (Ndu) are spoken inside it, while Skou (Sko) and Taiap (isolate) are spoken in relatively adjacent areas. Another oft-cited example is Alamblak (Bruce 1984; not in the sample), also a Sepik language of the same area, which has a system similar to that of Manambu (Aikhenvald 2008: 112).

Thus, gender assignment according to size and shape appears to be an areal feature, since it occurs in a wide area and in languages of different families. This

gives rise to an important question. Why would a system of gender assignment be areal when gender is such a stable and not easily borrowed feature? Although this is far beyond the scope of this study, there are some hints that this may be part of a larger cultural classificatory system (i.e., perceptual, not linguistic). The reason for such a possibility is that besides occurring in and around the Sepik area, there are other New Guinean languages where nouns are grouped based on size and shape with other nouns denoting male or female referents, even when there is no gender system. This is most apparent in the TNG languages of the central highlands; nouns in these languages can be categorized by the type of stance verb they occur with, so that males or large, long, or tall objects occur with ‘stand’, whereas women or small, short, or round objects occur with ‘sit’ (Foley 2000: 372). An example of such a language is Enga (Engan; New Guinea Highlands; not in the sample), which has seven different stance verbs, including *katengé* ‘stand’, which is used for referents considered tall, large, strong, and/or powerful such as ‘men’, ‘house’, and ‘tree’, and *pentengé* ‘sit’, which is used for referents considered small, squat, horizontal, and/or weak such as ‘woman’, ‘possum’, and ‘pond’ (Aikhenvald 2000: 158–159; Rumsey 2002). Thus, it appears that the perception of large, long, or tall objects being related to males and/or masculinity, and small, short, or round objects being related to females and/or femininity is a characteristic of New Guinea that extends beyond gender systems or the Sepik area.

## 5.2 Two separate systems of noun classification

In most gendered languages, gender constitutes a single system where each noun is assigned to a single class which is reflected in the form of indexation targets. However, there are also languages with two separate systems, both of which appear to constitute or be related to gender systems, but occur with different types of targets. Thus, in such a language each noun is assigned to not just one class, but to two different classes. In the sample of this study, five languages have such systems (see Table 15).

Even in the small sample of this study, the two separate systems range from languages with two more or less equally complex systems (i.e., with similar numbers of forms and uses) to languages where one system is more complex whereas the other is much less so. In order to retain the typological comparability of the results, a distinction has been made between systems of gender and systems of noun classifiers. However, it should be stated that there is a thin line between the two and they most certainly constitute two edges of the same continuum.

Table 15: Languages in the sample with separate gender and noun class systems

Separate systems	No. of lgs.	%	Languages
Yes	5	25%	Abau Burmese Mian Motuna Rotokas
No	15	75%	Ama Au Bukiyip Kuot Manambu Maybrat Mende Nalca Oksapmin Skou Taiap Teop Walman Warapu Yimas
Total:	20	100%	

Following these, four of the five languages with two systems of noun classification can be argued to exhibit one gender system and system of noun classifiers, whereas only Burmese has two systems which both satisfy the conditions for gender systems. In the first system, Burmese has three genders (masculine, feminine, and neuter), appearing as adjectival agreement suffixes (11a), which are further divided into two subgenders (animate and inanimate), each depending on the plural agreement marker (Donohue 2001: 105–106). However, in the second system (which Donohue calls a noun class system), Burmese has six genders (I–VI), which occur in verbal agreement prefixes (11b) (Donohue 2001: 101). In addition, there are three words which take both kinds of agreement: *-aysa* ‘one’, *-akasu* ‘all’, and *-asna* ‘white’ (11c).

(11) Burmeso (isolate) (Donohue 2001: 105, 109, 100)

- a. *Da de koya bek-**abo**.*  
1SG 1SG.POSS grandfather good-M.SG  
'My grandfather is well.'
- b. *Da mibo j-**ihi**-maru.*  
1SG banana V.SG-see-TPST  
'I saw a banana.'
- c. *Sunam **n**-asna-**b**.*  
axe.SG III.SG-white-M.SG  
'(The) axe is white.'

As expected from the number of genders being different, the two systems use different assignment rules. Both systems are sex-based with importance clearly put on sex and animacy, but none of them have only transparent semantic rules: e.g., 'wind' is neuter/III, 'rain' masculine/IV, and 'star' masculine/III (Donohue 2001: 103–107). A comparison of the overlap of the two systems is exemplified in Table 16, showing how members are assigned to both systems.

Near the other end of the spectrum lies Rotokas (North Bougainville). Rotokas has three genders, viz., masculine, feminine, and neuter, which appear e.g., in pronouns, demonstratives, adjectives, and verbs (12a) (Robinson 2011). However, Rotokas also has noun classifiers, which consist of two different sets. The first set consists of four classifiers; these distinguish between shape and size, and importantly occur on both attributive (12b) and predicative modifiers of the classified noun (Robinson 2011: 50).

(12) Rotokas (North Bougainville) (Robinson 2011: 149, 50)

- a. *Pita vaio ora Kariri ava-**si**-ei voka-sia*  
P. DL.ANIM and K. go-3DL.M-PRS walk-DEP.SEQ  
'Peter and Kariri are going for a walk.'
- b. *gorupasi **isi** rutu karuvera **isi** aio-a-voi*  
strong CL.**round** very Singapore CL.**round** eat-1SG-PRS  
'I am eating a really strong Singapore fruit.'

The other set of classifiers, which has more members and have collective meanings, occurs following, or instead of, the classified noun (Robinson 2011: 51). Interesting to note is that classified nouns become neuter in regards to gender agreement (Robinson 2011: 53).

Table 16: Comparison between genders and noun classes in Burmeso (adapted from Donohue 2001: 108). As in the source, the subgenders of the masculine, feminine, and neuter genders are shown.

Class	Masculine	Feminine	Neuter	M	F	N
				inanimate	inanimate	animate
I	male humans (most birds, animals etc.) 2SG PRO 1SG PRO	(birds of paradise) pigeons sago garden female humans	sea	neck	–	sea wound
II			nose ear eye	–	small goanna	string shapes sago rinser (lower)
III	axe papaya ground bird	–	(some small animals) bench	papaya rattan mountain lake (all tubers) upper sago trough female child	goanna	–
IV	head, flesh, feces, finger, elbow, sun, cloud, rain, sand, mud	–	–	–	–	–
V	–	–	(arrows)	banana	–	–
VI	–	–	–	coconut	–	–



Table 17: Numeral classifiers in Abau (adapted from Lock 2011:57)

Class	Characteristics	One	Two	Three
1	Human beings; spirits	<i>pru-eyn</i>	<i>pru-ey</i> s	<i>pru-ompri</i>
2	Non-human	<i>ka-mon</i>	<i>k-reys</i>	<i>k-rompri</i>
3	Small objects with some volume	<i>na-mon</i>	<i>na-reys</i>	<i>na-rompri</i>
4	Flat surface objects; experience nouns	<i>si-rom</i>	<i>s-ey</i> s	<i>s-ompri</i>
5	Long, relatively thin objects	<i>pi-ron</i>	<i>pi-reys</i>	<i>pi-rompri</i>
6	Geographical locations	<i>u-mon</i>	<i>u-reys</i>	<i>u-rompri</i>
7	Flat objects with hardly any volume	<i>i-mon</i>	<i>i-reys</i>	<i>i-rompri</i>
8	Certain type trees	<i>li-mon</i>	<i>li-reys</i>	<i>li-rompri</i>
9	Bundles of long non-cut items	<i>ein-mon</i>	<i>ein-deys</i>	<i>ein-rompri</i>
10	Temporal	<i>leik-mon</i>	<i>leik-reys</i>	<i>leik-rompri</i>
11	Bundles of long cut items	<i>hnaw-mon</i>	<i>hnaw-reys</i>	<i>hnaw-rompri</i>
12	Part of a long object	<i>houk-mon</i>	<i>houk-reys</i>	<i>houk-rompri</i>

Abau also exhibits a clear noun classifier system (Table 17). There are two genders in Abau, masculine and feminine, which follow opaque gender assignment rules and appear in e.g., pronouns and demonstratives. However, the numerals ‘one’, ‘two’, and ‘three’ do not agree with this system, but instead take one of twelve prefixes based on semantic criteria of the referent. However, the same noun can be used with different numeral classifiers in order to indicate a specific referent, so that e.g., *su piron* ‘one coconut’ refers to the whole coconut palm and not just the fruit, since class 5 signals long objects, while *su kamon* ‘one coconut’ is used when referring to just the fruit, since class 2 does not carry the semantic feature of length. It is thus evident that this system of noun classifiers is not lexically determined by the noun itself and thus not a gender system.

Mian has a similar albeit different system. In Mian, there is a set of verbal classificatory prefixes which are divided into six classes (Table 18). These prefixes are used only for around 50 verbs, the vast majority of which refer to forms of object manipulation, movement, and handling (Fedden 2011: 172). Once again, this is clearly not a full-fledged gender system, but rather a classifier system.

Table 18: Classifiers in Mian (adapted from Fedden 2011: 172)

Class	Characteristics	Verbal classificatory prefixes	
		Singular	Plural
1	Masculine	<i>do(b)-</i>	<i>do(l)-, dl-</i>
2	Feminine	<i>om-</i>	
3	Long object	<i>to(b)-</i>	<i>tebe(l)-</i>
4	Bundle-like object	<i>go(l)-</i>	<i>gule(l)-</i>
5	Flat object	<i>gam-</i>	<i>geme(l)-</i>
6	Residue class	<i>o(b)-</i>	<i>o(l)-</i>

Finally, Motuna is a particularly interesting case since its secondary system lies near the boundary between genders and noun classifiers. Besides its gender system (described in §3.3), Motuna has another noun classification system consisting of 51 different classifiers, which are visible in the forms of adjectives, verbs, participial clauses, articles, demonstratives, possessive pronouns, and numerals (Onishi 1994: 162–163). Thus, as for indexation, the system is very reminiscent of a gender system. However, the classes are not lexically determined, meaning that the same noun may occur with various classifiers depending on the referent. Furthermore, as expected for a noun classifier system, the classifiers refer properties such as size, shape, type of vegetable, and collectives (e.g., ‘bundle’, ‘packet’). Thus, *moo* ‘coconut’ can occur with classes 4 *-mung* ‘plant/fruit/nut/egg/things made of plant/coin’ (> ‘coconut (nut/tree)'), 5 *-ri* ‘nut with hard shell’ (> ‘coconut’), 6 *-mo* ‘bunch of nuts’ (> ‘coconut’), 13 *-ri* ‘round object’ (> ‘coconut’), and 30 *-ita* ‘half/side’ (> ‘half coconut shell’) (Onishi 1994: 166–167). Therefore, this system in Motuna is a system of noun classifiers, not genders.

Despite the small size of the sample used in this study, the proportion and the geographic and genealogical spread of languages with two separate systems of nominal classification indicate that the phenomenon is rather common and widespread in New Guinea. Besides the languages of this study, two of which are mentioned by Foley (2000: 373), viz., Burmeso and Motuna, similar systems have been noted in the Sepik languages Iwam, Wogamusin, and Chenapien, which together with their relative Abau (which is included in this sample) suggest that this is a feature of the Sepik family (Lock 2011: 46). However, it does not appear to be common outside of New Guinea, as systems of this type only occur in a few Indic, Dravidian, Iranian, and some Arawak languages (Aikhenvald 2008: 185).

### 5.3 No gender distinctions in pronouns

According to Greenberg's (1963: 90) 43<sup>rd</sup> Universal, "[if] a language has gender categories in the noun, it has gender categories in the pronoun".<sup>15</sup> However, this generalization is not reflected in the languages sampled for this study, where four languages do not exhibit gender in pronouns (see Table 19).<sup>16</sup>

Table 19: Occurrence of gender distinctions in independent pronouns in the sample

Gender in pronouns	No. of lgs.	%	Languages
Yes	16	80%	Abau Au Bukiyip Kuot Manambu Maybrat Mende Mian Motuna Oksapmin Rotokas Skou Taiap Walman Warapu Yimas
No	4	20%	Ama Burmese Nalca Teop
Total:	20	100%	

<sup>15</sup>'Pronoun' is here understood as 'independent pronoun'.

<sup>16</sup> As in §3.4, the demonstratives in Kuot and Yimas with pronominal functions are here understood as pronouns for the purpose of typological comparison, just as the present author would do for the Latin *is*, *ea*, and *id*, regardless of the proper language-internal analysis. Nevertheless, if they should rather not be regarded as pronouns, the point of this section would be even stronger.

As seen in the above table, almost a quarter of the languages in the sample have no gender distinctions in independent pronouns. In comparison, only two languages (Mende and Menya) have gender distinctions solely in pronouns.

While these results are interesting, the phenomenon can be found in other languages as well. This can be investigated by comparing two *WALS* chapters, viz., Corbett's (2013a) chapter on number of genders and Siewierska's (2013) chapter on gender distinctions in independent pronouns. These chapters do not share the same sample: Corbett's sample consists of 257 languages, whereas Siewierska's contains 378 languages. Of these languages, 188 occur in both samples, 74 of which have gender systems. Of these remaining 74 gendered languages (which of course should not be assumed to be representative of anything), surprisingly, 15 languages (20%) do not show gender distinctions in independent pronouns. Coincidentally, this is the same ratio as in New Guinea as shown in Table 18 above. Thus, it is clear that Greenberg's statement is not universal, although it certainly is a common pattern.

#### 5.4 Gender indexation on verbs

According to Greenberg's 31<sup>st</sup> Universal, "if either the subject or object noun agrees with the verb in gender, then the adjective always agrees with the noun in gender." That is, if the verbs are indexing targets, so are adjectives. However, this generalization is not reflected in the distribution of values of the fourth classification criteria in the languages sampled for this study (see Table 8). Three of the 15 languages with gender marking on verbs show no indexation on adjectives.

The results are even more striking when compared with Bybee (1985). In her survey of fifty languages, only 16% of the languages showed gender in verbs (Bybee 1985: 18). However, in the sample of this survey, 75% of the languages have gender marking on verbs, with Ama even having it as the only indexing target. Verbs thus seem to be more prototypical indexing targets than adjectives in the sample of this study, and it would be interesting to conduct further studies on this with a larger and worldwide sample.

### 6 Conclusions and further studies

The languages of New Guinea show remarkable diversity in grammatical gender, but there are still common patterns. Except Teop (Austronesian, Oceanic), all languages in the sample have sex-based gender systems. More than half of the languages have only two genders, and only Bukiyip (Torricelli) and Yimas (Lower

Sepik) have very large systems, with 18 and around a dozen genders respectively. In the vast majority of the languages, gender assignment is semantic. Half of the languages have four or more indexing targets, most commonly pronouns and verbs. Gender marking on nouns is rare and occurs in only three languages in the sample. The typological comparison suggests that the genders systems of New Guinea are remarkably representative of the world. Sex-based gender systems are more common in both New Guinea and the world, and the ratio of numbers of genders are very similar, with the rate of occurrence of the values being two > three ≥ five or more > four genders. Semantic and formal gender assignment occurs in slightly more than half of the languages of the world, while it is much more uncommon in New Guinea. The gender systems of New Guinea and Africa are very different. This depends largely on the numerous Bantu languages, which make the languages of Africa whole less diverse than the sample of this study. The most significant difference is the prevalence of non-sex-based gender systems and gender marking on nouns in Africa, whereas the opposite is true in New Guinea. This suggests that they may have different diachronical origins.

Four special characteristics have been found in the gender systems of New Guinea, none of which are typologically common. Firstly, four languages of the sample share the property of size and shape as important criteria for gender assignment. In these languages, nouns denoting large and/or long objects are masculine, whereas small and/or short items are feminine. This characteristic is also shared with many African languages. Secondly, five languages of the sample have two separate nominal classification systems. In these languages, each noun is assigned to two classes which are reflected in different indexing targets, although only Burmeso exhibits two equivalent gender systems whereas the others rather distinguish between genders and noun classifiers. Thirdly, four languages in the sample have no gender distinctions in pronouns, which is unexpected according to Greenberg's 43<sup>rd</sup> Universal. Finally, verbs are the most common gender-indexing targets in the languages of the sample, which is uncommon. In three languages of the sample, verbs are indexing targets while adjectives are not, which contradicts Greenberg's 31<sup>st</sup> Universal.

Future studies should consider more languages and be proportional, as well as aim at investigating how the gender systems of New Guinea may affect the theory of gender. There are also more specific areas of study that would benefit from further research. Firstly, the special characteristics discussed in this study could benefit from more research. One example is gender assignment based on size and shape, which appears to be a feature of the Sepik area. However, Skou (Sko) and Taiap (isolate) are spoken outside of the immediate area, and similar distinctions

have been found in non-gendered languages of New Guinea. It would thus be interesting to investigate the actual geographical distribution of such systems. Also, the inclusion of the criterion of manipulability of gender assignment as used in Di Garbo (2014) would probably further improve the comparison between gender in New Guinea with Africa.

It would also be interesting to investigate features not discussed in this study. One such feature is pluralia tantum, i.e., plural nouns with no or only an unusual singular form (Koptjevskaja-Tamm & Wälchli 2001: 629), for which there are indications that it may be relevant for gender. This can be seen in Ama (Left May), which has a separate compound gender containing nouns denoting referents with many parts, e.g., heaps, piles, and mass nouns (Årsjö 1999: 68). For a discussion of pluralia tantum in languages of New Guinea see also Olsson (2019 [this volume]) and Dryer (2019 [this volume]).

Future studies could also investigate the diachrony of gender in New Guinea. Some languages of New Guinea have been found to have diachronically young gender systems, including Nalca (TNG, Mek) of the sample of the present study, and the prevalence of sex-based systems suggest that many gender systems in New Guinea have diachronic origins different from e.g., the non-sex-based gender systems of Africa.

## Special abbreviations

The following abbreviations are not found in the Leipzig Glossing Rules:

I, II, III etc.	gender I, II, III etc.	INTEN	intensifier
ANIM	animate	N-	non-
C	common gender	NNOUN	non-noun gender
CL	classifier	PRO	pronoun
CO1	common gender 1	RED	reduplication
DEFAULT	default gender	SEQ	sequential
DEP	dependent (verb)	TPST	today's past/hodiernal past
DL	dual	U	unmarked gender
HAB	habitual		

## References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.

- Aikhenvald, Alexandra Y. 2008. *The Manambu language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.
- Årsjö, Britten. 1999. *Words in Ama*. Uppsala: Uppsala University MA thesis.
- Bakker, Dik. 2012. Language sampling. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology* (Oxford Handbooks Online). Oxford: Oxford University Press. <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199281251.001.0001/oxfordhb-9780199281251>, accessed 2015-5-15.
- Brown, Lea & Matthew S. Dryer. 2008. Diminutive as an inflectional category in Walman. <http://linguistics.buffalo.edu/people/faculty/dryer/dryer/BrownDryerWalmanDimin.pdf>, accessed 2015-2-4. Unpublished manuscript.
- Bruce, Leslie P. 1984. *The Alamblak language of Papua New Guinea (East Sepik)* (Pacific Linguistics C 81). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- Bybee, Joan. 1985. Diagrammatic iconicity in stem-inflection relations. In John Haiman (ed.), *Iconicity in syntax*, 11–48. Amsterdam: John Benjamins.
- Carter, Hazel. 2002. *An outline of Chitonga grammar*. Lousaka: Bookwords Publishers.
- Conrad, Robert J. & Kepas Wogiga. 1991. *An outline of Bukiyip grammar* (Pacific Linguistics C 113). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2013a. Number of genders. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/30>.
- Corbett, Greville G. 2013b. Sex-based and non-sex-based gender systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/31>.
- Corbett, Greville G. 2013c. 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. <http://wals.info/chapter/32>.
- Corris, Miriam. 2005. *A grammar of Barupu: A language of Papua New Guinea*. Sydney: University of Sydney dissertation.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William. 2003. *Typology and universals*. 2nd edn. 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 Siewierska*, 95–118. Berlin: De Gruyter.
- Di Garbo, Francesca. 2014. *Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of Africa*. Stockholm University dissertation.
- Dol, Philomena Hedwig. 2007. *A grammar of Maybrat: A language of the Bird's Head Peninsula, Papua Province, Indonesia*. Canberra: Pacific Linguistics.
- Donohue, Mark. 2001. Animacy, class and gender in Burmeso. In Andrew Pawley, Malcolm Ross & Darrell Tryon (eds.), *The boy from Bundaberg: Studies in Melanesian linguistics in honour of Tom Dutton* (Pacific Linguistics 514), 97–115. Canberra: Australian National University.
- Donohue, Mark. 2004. A grammar of the Skou language of New Guinea. [http://pubman.mpdl.mpg.de/pubman/item/escidoc:402710/component/escidoc:402709/skou%5C\\_donohue2004%5C\\_s.pdf](http://pubman.mpdl.mpg.de/pubman/item/escidoc:402710/component/escidoc:402709/skou%5C_donohue2004%5C_s.pdf), accessed 2015-2-6. Unpublished manuscript.
- Dryer, Matthew S. 2019. Gender in Walman. In Francesca Di Garbo, Bruno Olson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 171–196. Berlin: Language Science Press. DOI:??
- Fedden, Sebastian. 2011. *A grammar of Mian*. Berlin: Mouton de Gruyter.
- Fedden, Sebastian & Greville G. Corbett. 2017. Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa: A journal of general linguistics* 2(1). 1–47. DOI:10.5334/gjgl.177
- Foley, William A. 1991. *The Yimas language of New Guinea*. Stanford: Stanford University Press.
- Foley, William A. 2000. The languages of New Guinea. *Annual Review of Anthropology* 29. 357–404.
- Frowein, Friedel Martin. 2011. *A grammar of Siar: An Oceanic language of New Ireland province, Papua New Guinea*. Victoria: La Trobe University dissertation.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–113. Cambridge: MIT Press.
- Heine, Bernd. 1982. African noun class systems. In Hansjakob Seiler & Christian Lehmann (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen, Teil I: Bereich und Ordnung der Phänomene*, 189–216. Tübingen: Narr.
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.



- Hoel, Hanna Marie, Tarja Ikäheimonen & Michiyo Nozawa. 1994. Mende grammar essentials. [http://www-01.sil.org/pacific/png/pubs/928474531144/Mende\\_Grammar\\_Essentials.pdf](http://www-01.sil.org/pacific/png/pubs/928474531144/Mende_Grammar_Essentials.pdf), accessed 2015-3-20. Unpublished manuscript.
- Iemmolo, Giorgio. 2011. *Towards a typological study of differential object marking and differential object indexation*. Pavia: Università degli studi di Pavia dissertation.
- Koptjevskaja-Tamm, Maria & Bernhard Wälchli. 2001. The Circum-Baltic languages: An areal-typological approach. In Östen Dahl & Maria Koptjevskaja-Tamm (eds.), *Circum-Baltic languages: Typology and contact*, vol. 2, 615–750. Amsterdam: John Benjamins.
- Kulick, Don & Christopher Stroud. 1992. The structure of the Taiap (Gapun) language. In Thomas E. Dutton, Malcolm Ross & Darrell T. Tryon (eds.), *The language game: Papers in memory of Donald C. Laycock* (Pacific Linguistics C 110), 203–226. Canberra: Australian National University.
- Lindström, Eva. 2002. *Topics in the grammar of Kuot: A non-Austronesian language of New Ireland, Papua New Guinea*. Stockholm: Stockholm University dissertation.
- Lock, Arjen. 2011. *Abau grammar* (Data Papers on Papua New Guinea Languages 57). Ukarumpa: SIL-PNG Academic Publications.
- Loughnane, Robyn. 2009. *A grammar of Oksapmin*. University of Melbourne dissertation.
- Luraghi, Silvia. 2011. The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45(2). 435–464.
- Mosel, Ulrike & Ruth Spriggs. 2000. Gender in Teop (Bougainville, Papua New Guinea). In Barbara Unterbeck (ed.), *Gender in grammar and cognition I: Approaches to gender* (Trends in Linguistics. Studies and Monographs 124), 321–349. Berlin: Mouton de Gruyter.
- Olsson, Bruno. 2019. The gender system of Coastal Marind. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 197–223. Berlin: Language Science Press. DOI:??
- Onishi, Masayuki. 1994. *A grammar of Motuna (Bougainville, Papua New Guinea)*. Canberra: Australian National University dissertation.
- Phillips, Colin. 1993. Conditions on agreement in Yimas. In Jonathan D. Bobaljik & Colin Phillips (eds.), *Papers on case & agreement I* (MIT Working Papers in Linguistics 18), 173–213. Cambridge: MIT.
- Robinson, Stuart Payton. 2011. *Split intransitivity in Rotokas, a Papuan language of Bougainville*. Nijmegen: Radboud Universiteit dissertation.

- Rumsey, Alan. 2002. On the grammaticalization of posture verbs in Papuan languages, its bodily basis and cultural correlates. In John Newman (ed.), *The linguistics of sitting, standing and lying*, 179–211. Amsterdam: John Benjamins.
- Scorza, David. 1985. A sketch of Au morphology and syntax. In *Papers in New Guinea linguistics* 22 (Pacific Linguistics A 63), 215–273. Canberra: Australian National University.
- Siewierska, Anna. 2013. Gender distinctions in independent personal pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/44>.
- Steele, Susan. 1978. Word order variation: A typological study. In Joseph H. Greenberg, Charles A. Ferguson & Edith A. Moravcsik (eds.), *Universals of human language*, vol. 4: *Syntax*, 585–624. Stanford: Stanford University Press.
- Svärd, Erik. 2013. Selected topics in the grammar of Nalca. <http://www.diva-portal.org/smash/record.jsf?searchId=1%5C&%20pid=diva2:631006>, accessed 2013-11-9. BA-thesis, Stockholm University.
- Svärd, Erik. 2015. *Grammatical gender in New Guinea*. Stockholm: Stockholm University MA thesis.
- Tharp, Doug. 1996. Sulka grammar essentials. In John M. Clifton (ed.), *Two non-Austronesian grammars from the islands*, 77–179. Ukarumpa: Summer Institute of Linguistics.
- Wälchli, Bernhard. 2018. The rise of gender in Nalca (Mek, Tanah Papua): The drift towards the canonical gender attractor. In Sebastian Fedden, Jenny Audring & Greville G. Corbett (eds.), *Non-canonical gender systems*, 68–99. Oxford: Oxford University Press.

## **Part IV**

# **South Asia**



## Chapter 10

# Gender typology and gender (in)stability in Hindu Kush Indo-Aryan languages

Henrik Liljegren

Stockholm University

This paper investigates the phenomenon of gender as it appears in 25 Indo-Aryan languages (sometimes referred to as “Dardic”) spoken in the Hindu Kush-Karakorum region – the mountainous areas of northeastern Afghanistan, northern Pakistan and the disputed territory of Kashmir. Looking at each language in terms of the number of genders present, to what extent these are sex-based or non-sex-based, how gender relates to declensional differences, and what systems of assignment are applied, we arrive at a micro-typology of gender in Hindu Kush Indo-Aryan, including a characterization of these systems in terms of their general complexity. Considering the relatively close genealogical ties, the languages display a number of unexpected and significant differences. While the inherited sex-based gender system is clearly preserved in most of the languages, and perhaps even strengthened in some, it is curiously missing altogether in others (such as in Kalasha and Khovar) or seems to be subject to considerable erosion (e.g. in Dameli). That the languages of the latter kind are all found at the northwestern outskirts of the Indo-Aryan world suggests non-trivial interaction with neighbouring languages without gender or with markedly different assignment systems. In terms of complexity, the southwestern-most corner of the region stands out; here we find a few languages (primarily belonging to the Pashai group) that combine inherited sex-based gender differentiation with animacy-related distinctions resulting in highly complex agreement patterns. The findings are discussed in the light of earlier observations of linguistic areality or substratal influence in the region, involving Indo-Aryan, Iranian, Nuristani, Tibeto-Burman, Turkic languages and Burushaski. The present study draws from the analysis of earlier publications as well as from entirely novel field data.

**Keywords:** Afghanistan, animacy, complexity, Dardic, gender pervasiveness, Indo-Aryan, Kashmir, non-sex-based gender, Pakistan, sex-based gender.



Henrik Liljegren. 2019. Gender typology and gender (in)stability in Hindu Kush Indo-Aryan languages. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 279–328. Berlin: Language Science Press. DOI:??

## 1 Introduction

At the very northern fringe of the Indo-Aryan world (approximately what lies north of the 34<sup>th</sup> parallel) we find a group of languages that historically and culturally are somewhat outside the sphere of the main Indo-Aryan languages of the subcontinent (Masica 1991: 20–21). Geographically, this group is wedged in between Iranian on its western side and Tibeto-Burman on its eastern side, and the distance to the Turkic belt of Central Asia is negligible at its farthest extension, even if it is not immediately adjacent. This extremely mountainous and multilingual region (see Figure 1), lies where the territories of Afghanistan, Pakistan and India-administered Kashmir meet. Henceforth, I will refer to this region as the Hindu Kush.<sup>1</sup> Apart from the languages and genera already mentioned, this region is also home to Nuristani – a third, but numerically small, branch of Indo-Iranian (Strand 1973: 297–298) – and to the isolate Burushaski.

The languages in question have been subject to a great deal of debate as to whether they are truly Indo-Aryan, constitute a genealogical unit of their own, or represent (perhaps along with the Nuristani languages) a transitional group between Indo-Aryan and Iranian. A term frequently used collectively for these languages is “Dardic”. However, few modern linguists use this term as anything else than a convenient umbrella term for a group of languages that are characterized – but not equally so – by a few salient retentions from previous stages of Indo-Aryan (Morgenstierne 1974: 3), but also have some contact-related developments in common (Bashir 2003: 821–822). Contact in that case includes mutual contact between the various Indo-Aryan linguistic communities as well as significant contact with adjacent communities belonging to other genera (Liljegren 2017). This non-committal line is also taken here regarding this grouping, but in order to avoid a stronger interpretation of “Dardic” than warranted, the term is abandoned in favour of Hindu Kush Indo-Aryan (HKIA) (Liljegren 2014: 135; Heegård Petersen 2015: 23), again without any claim of classificatory significance in the traditional sense. While the region for quite some time has been identified as particularly interesting in terms of areality and language contact (Emeneau 1965; Skalmowski 1985; Masica 1991: 43; Masica 2001: 259), and a number of features have been suggested as characteristic (Bashir 1988: 392–420; Bashir 1996; Bashir 2003: 821–823; Èdel’man 1980; Èdel’man 1983: 35–59; Fussman 1972: 389–399; Tikkanen 1999; 2008; Baart 2014; Toporov 1970), relatively little detailed and systematic areal-linguistic research has been carried out so far.

---

<sup>1</sup>Strictly speaking, this region only partly overlaps with the Hindu Kush mountain range, while also overlapping with the Karakorum and the westernmost extension of the Himalayas.

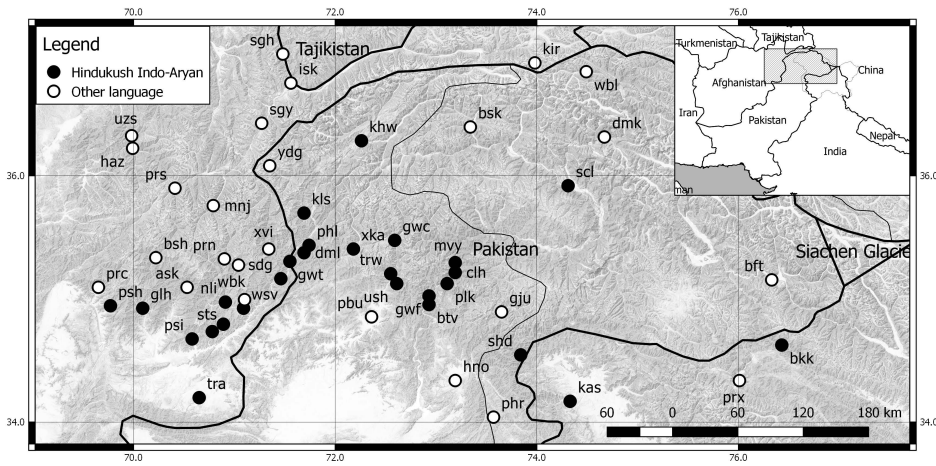


Figure 1: The Hindu Kush-Karakoram region with languages plotted (see Table 1 for an explanation of the 3-letter codes)

Regarding the ancestral nominal system, evidenced in Old Indo-Aryan as well as in Middle Indo-Aryan, it encompassed three gender values: masculine, feminine and neuter. In the Indo-Aryan world in general, these three values are only preserved in the modern languages in the southern part of the subcontinent, whereas a simplified two-value system (masculine vs. feminine, mainly as a result of neuter collapsing with masculine) dominates the large central and western parts. Such distinctions have altogether vanished in the northeast (Masica 1991: 217–223). The somewhat unexpected distribution and display of grammatical gender in the languages at the northern and western frontier of Indo-Aryan (viz. the Hindu Kush) was pointed out by Emeneau (1965: 68–71) half a century ago, but apart from Morgenstierne’s (1950: 19–20) tabulation, no systematic attempt has to my knowledge been made to account for gender distribution and manifestation across HKIA. This study tries to rectify that by showing the results of a survey of the following gender-related features – partly inspired by a number of contributions to the *World atlas of language structures (WALS)* – for each HKIA language for which there is data:

- The presence and number of gender categories (as evidenced by agreement patterns), and their basis, whether sex-based or non-sex-based (Corbett 2013a,b).
- The pervasiveness of gender, i.e. how gender is manifested in each language system in terms of (the types and numbers of) indexed domains.

- The assignment criteria at work: whether semantic or semantic-formal (Corbett 2013c).
- The presence and manifestation of pronominal gender (Siewierska 2013).

In the process of discussing and summarising these results, particularly in terms of the relative complexity of these systems, and in the light of areal patterning, a micro-typology of gender in HKIA emerges:

- The inherited sex-based system is largely preserved, but has disappeared in two of the languages at the Northwestern fringe of the Hindu Kush and is possibly eroding in a few other languages spoken in the same part of the region.
- An animacy-based system (almost exclusively marked on copulas or copula-based verbal categories) characterizes a number of the western-most languages of the region. In some cases it co-exists with a sex-based system; in others it occurs instead of a sex-based system or has contributed to a restructuring of the system as a whole.
- Gender is deeply entrenched (reflected in more target domains) in the East, i.e. in the languages spoken in areas contiguous with the main Indo-Aryan belt, whereas such pervasiveness is fading out toward the West.
- The results also suggest a weaker tendency toward semantic transparency in the gender systems in the North and a reinforcement of formal assignment, along with object agreement, in the South.

## **2 Hindu Kush Indo-Aryan and other languages in the region**

Today, there are 28 distinct HKIA languages, i.e. languages identified as “Dardic” by the language catalogue Ethnologue (Eberhard et al. 2019), spoken in the region, the great majority of them on Pakistani soil or in areas of Kashmir now under Pakistani control. At least six clusters of related languages can be identified, mainly going with Bashir (2003: 824–825) and the classification used in Glottolog (Hammarström et al. 2018), although the definitive placement of a few of the individual languages is still pending (Dameli, Tirahi and Wotapuri-Katarqalai). All HKIA languages are presented in Table 1, roughly according to their geographical distribution, from west to east in a crescent-like fashion (see Figure 1). No



attempt has been made here to represent relatedness below the level of these six groupings.

Some of these groupings are tighter, i.e. internally less diverse, than others. This is one reason why they sometimes are treated as single languages with a number of dialects rather than as groupings of separate languages. That especially applies to Kashmiri, Shina and Pashai. The relatedness between the two Chitral group languages, Khowar and Kalasha, is also apparent from a number of features that single these two out from the rest of HKIA. The latter two were assumed by Morgenstierne (1932: 51) to represent the first wave of Indo-Aryan settlers moving in from the lowlands in the South.

If we, for the sake of simplicity, define the Hindu Kush region as the window between the longitudes 34 and 37 N and the latitudes 69 and 77 E, another 25 languages are spoken here. At least four other languages (or continua), traditionally described as belonging to sub-branches of Indo-Aryan with their geographical centres outside of the Hindu Kush region, are also found in the Hindu Kush region, or their geographical extension overlaps to a considerable extent with it: Hindko [hno], Pahari-Pothwari [phr], Gojri [gju] and Domaaki [dmk]. Hindko and Pahari-Pothwari are essentially part of a Punjabi macro-language extended far beyond the region, and as such they represent the closest main Indo-Aryan neighbour of HKIA. Gojri is the language of nomadic or semi-nomadic Gujurs, spoken in pockets throughout the region and beyond. The closest linguistic relatives of Rajasthani Indo-Aryan Gojri is, however, to be found at a considerable distance from the present region, deep into the main belt of Indo-Aryan. The closest relatives of Domaaki are likewise to be found in the plains of North India. Domaaki, however, is interesting from an areal point of view; as the language of a small enclave of musicians and blacksmiths surrounded by locally dominant speaker groups of Shina and Burushaski, it has during its 200–300 years in the area acquired a number of features typical of HKIA (Weinreich 2011: 165–166).

A number of the surrounding languages in the West are Iranian. Pashto [pbu] and Dari [prs], the two representing two completely different branches of Iranian, are both important lingua francas in parts of the region and well beyond. Dari is essentially the standard or literary type of Eastern Persian used in Afghanistan, while various names occur in reference to regional or local varieties, such as Tajik in north-eastern Afghanistan and neighbouring Tajikistan. Some of those may very well be considered languages in their own rights, e.g. Hazaragi [haz]. Most of the other Iranian languages (all very distantly related to either Pashto or Dari) are relatively minor, with a local scope only; in Afghanistan, Parachi [prc], Munji [mnj], Sanglechi [sgy], Ishkashimi [isk] and Shughni [sgh]; in Pakistan, Yidgha [ydg], basically a dialect of the same language as Munji; in Pakistan and

Table 1: Hindu Kush Indo-Aryan languages (with 3-letter ISO codes and the areas and countries where they are spoken), arranged in sub-groupings

Group	Language	code	Area (Country)
Pashai	Northwest Pashai	[glh]	Kabul, Kapisa, Konar, Laghman, Nurestan (Afg)
	Southwest Pashai	[psh]	Kabul, Kapisa (Afg)
	Southeast Pashai	[psi]	Nangarhar, Laghman (Afg)
	Northeast Pashai	[aee]	Konar, Nangarhar (Afg)
Kunar	Shumashti	[sts]	Konar (Afg)
	Grangali	[nli]	Konar, Nangarhar (Afg)
	Gawarbat	[gwt]	Konar (Afg), Chitral (Pak)
	Dameli	[dml]	Chitral (Pak)
Chitral	Kalasha	[kls]	Chitral (Pak)
	Khowar	[khw]	Chitral, Gilgit-Baltistan (Pak)
Kohistani	Tirahi	[tra]	Nangarhar (Afg)
	Wotapuri-Katarqalai	[wsv]	Nurestan (Afg)
	Gawri (Kalami)	[gwc]	Upper Dir, Swat (Pak)
	Torwali	[trw]	Swat (Pak)
	Indus Kohistani	[mvy]	Kohistan (Pak)
	Gowro	[gwf]	Kohistan (Pak)
	Chilisso	[clh]	Kohistan (Pak)
	Bateri	[btv]	Kohistan (Pak)
	Mankiyali	[nlm]	Mansehra (Pak)
	Sawi	[sdg]	Konar (Afg)
Shina	Palula	[phl]	Chitral (Pak)
	Kalkoti	[xka]	Upper Dir (Pak)
	Ushojo	[ush]	Swat (Pak)
	Kohistani Shina	[plk]	Kohistan (Pak)
	Kundal Shahi	[shd]	Jammu & Kashmir (Pak)
	Shina (Gilgiti)	[scl]	Gilgit-Baltistan (Pak), Jammu & Kashmir (Ind)
	Brokskat	[bkk]	Jammu & Kashmir (Ind)
	Standard Kashmiri	[kas]	Jammu & Kashmir (Ind), Jammu & Kashmir (Pak)

Afghanistan as well as in adjacent areas of Tajikistan and China, Wakhi [wbl] is spoken.

All of the five to six Nuristani languages are spoken in a geographically confined area in Afghanistan's Nurestan Province, close to the Pakistan border (with some spill-over into adjacent Chitral): Kati [bsh], Kamviri [xvi] (more correctly a dialect rather than a separate language from the aforementioned), Waigali [wbk], Ashkun [ask], Tregami [trm] and Prasun [prn]. Two Turkic languages are spoken at the northern periphery of the region: Uzbek [uzs] and Kirghiz [kir]; and in the East two with each other closely related Tibeto-Burman languages are found: Balti [bft] and Purik [prx]. The already-mentioned language isolate Burushaski is spoken in the extreme North of Pakistan's Gilgit-Baltistan region.

### 3 Sample and data

The sparsity of data points in large-scale typological enterprises such as *WALS* stresses the need for different selectional criteria when it comes to areal-typological or micro-typological studies. For instance, three of the *WALS* features (30A, 31A, 32A) that deal with gender include in their 257-language sample only five of the languages spoken in the Hindu Kush (Burushaski, Kashmiri, Kirghiz, Pashto and Uzbek), and of them only one (Kashmiri) is a HKIA language (Corbett 2013a,b,c). For the feature surveying pronominal gender (44A), the corresponding figures are 2 (Burushaski and Kashmiri) and 1 (Kashmiri), respectively, in a world-wide 378-language sample (Siewierska 2013).

It was therefore the aim of this survey to draw data from as many as possible of the 28 above-mentioned HKIA languages, rather than trying to identify and justify a smaller sample. This posed some challenges, as the quality and amount of documentation vary greatly from language to language. However, by combining available published descriptions with my own field data from a variety of languages in the region, it has been possible to find out which are the main characteristics and values (as presented in §1) for as many as 25 of them. I saw a definite need to exclude Gowro, Chilisso and Mankiyali due to lack of adequate data, but this should probably not distort the overall picture in any significant way, since the preliminary analysis shows that at least Gowro and Chilisso are relatively closely linked to Indus Kohistani (Bashir 2003: 874). The addition of unpublished field data was particularly important concerning the under-researched languages Bateri, Kalkoti and Ushojo. In Table 2, the sources of information for each language are specified.

Table 2: Data sources for Hindu Kush Indo-Aryan

Language	Sources
Northwest Pashai	(Morgenstierne 1967: 143–203); own data
Southwest Pashai	(Morgenstierne 1967: 45–142)
Southeast Pashai	(Morgenstierne 1967: 251–297; Lehr 2014); own data
Northeast Pashai	(Morgenstierne 1967: 205–249); own data
Shumashti	(Morgenstierne 1945)
Grangali	(Bashir 2003: 837–839; Grjunberg 1971)
Gawarbati	(Morgenstierne 1950); own data
Dameli	(Morgenstierne 1942; Perder 2013); own data
Kalasha	(Heegård Petersen 2015: 35–49; Bashir 1988); own data
Khowar	(Bashir 2003: 844–849); own data
Tirahi	(Morgenstierne 1934b; Grierson 1927: 265–327)
Wotapuri-Katarqalai	(Buddruss 1960)
Gawri (Kalami)	(Baart 1997; 1999); own data
Torwali	(Lunsford 2001; Bashir 2003: 864–869; Grierson 1929); own data
Indus Kohistani	(Hallberg & Hallberg 1999; Bashir 2003: 874–877; Lubberger 2014); own data
Bateri	(Hallberg & O’Leary 1992: 207–225, 249–251); own data
Sawi	(Buddruss 1967; Liljegren 2009: 43–48); own data
Palula	(Liljegren 2016); own data
Kalkoti	(Liljegren 2009: 43–48; Liljegren 2013); own data
Ushojo	(Decker 1992); own data
Kohistani Shina	(Schmidt & Kohistani 2008); own data
Kundal Shahi	(Rehman & Baart 2005); own data
Shina (Gilgiti)	(Bailey 1924; Degener 2008: 13–65; Radloff & Shakil 1998: 183–192); own data
Brokskat	(Ramaswami 1982; Sharma 1998)
Standard Kashmiri	(Koul 2003; Verbeke 2013: 175–211); own data

## 4 Gender Categories and their basis

The first question to address is whether gender a distinctive feature; and, if it is, also how many genders there are in the language. Here I align myself with the view that membership in a particular gender category in contrast with one or

more other such categories in the language in question is inherent to a noun but has to be evidenced by grammatical contrasts outside the noun itself, for instance in the form of adjectival or verbal agreement (Corbett 2014: 89–90; Hockett 1958: 231–233; Greenberg 1978: 50). Another relevant question is whether the gender system is based on, or primarily linked to, biological sex, or to something other than sex. Surveying the languages in our sample, we find (Table 3) that all of them display gender distinctions, one way or the other, with the possible exception of some dialects of NW Pashai.<sup>2</sup>

As can also be seen in Table 3, the basis for such distinctions is not the same for all of the languages. In the great majority of the languages (23 out of 25), the gender system, as it is mirrored in agreement, is clearly sex-based, having (at least) a two-way, female vs. male, differentiation at its core (as in many other Indo-Aryan languages in general). This is seen in example (1) from Ushojo, where ‘boy’ in (a) triggers masculine verb agreement, and ‘girl’ in (b) triggers feminine agreement. This masculine–feminine differentiation also extends into the inanimate realm: ‘wind’, in (c), is assigned feminine gender, and ‘coldness’, in (d), is assigned masculine gender.

(1) Ushojo (Own data)

- a. *ek phoó asíl-u, se seekel-aá yáa áal-u*  
 one boy(M) be.PST-M.SG 3SG.NOM bicycle-LOC going come.PFV-M.SG  
 ‘There was a boy, he came riding on a bicycle.’  
 (USH-PearStoryAH:001)
- b. *ek phuí ... seekal-aá yáa mušiin tarapayá áal-i*  
 one girl(F) bicycle-LOC going to.near in.direction come.PFV-F.SG  
 ‘A girl... came in his direction, riding on a bicycle.’  
 (USH-PearStoryAH:012)
- c. *axeér oóš čóku bíl-i*  
 finally wind(F) quiet become.PFV-F.SG  
 ‘Finally the wind gave up.’ (USH-NorthwindAH:007)
- d. *maáti šídal bíl-u*  
 1SG.DAT coldness(M) become.PFV-M.SG  
 ‘I feel cold [lit. Coldness came to me].’ (USH-ValQuestAH:060)

<sup>2</sup>The preliminary analysis of my own data, from three NW Pashai locations (Sanjan, Alasai and Alishang) indicates the overall presence of sex-based adjectival gender agreement, whereas clear evidence of animacy-based differentiation is lacking in these particular varieties. While those findings have guided the present treatment, Morgenstierne’s (1967: 150–151, 173–176) study suggests a great deal of dialectal variation within NW Pashai as far as the presence/absence of both sex-based and animacy-based gender are concerned.

Table 3: The presence of gender (sex-based, non-sex-based) in Hindu Kush Indo-Aryan

Language	Number of genders	Sex-based gender	Non-sex-based gender
Southwest Pashai	4	✓	✓
Southeast Pashai	4	✓	✓
Northeast Pashai	4	✓	✓
Shumashti	3–4	✓	✓
Dameli	3	✓	✓
Kalasha	2		✓
Khowar	2		✓
Northwest Pashai	2	✓	
Grangali	2	✓	
Gawarbatl	2	✓	
Tirahi	2	✓	
Wotapuri-Katarqalai	2	✓	
Gawri (Kalami)	2	✓	
Torwali	2	✓	
Indus Kohistani	2	✓	
Bateri	2	✓	
Sawi	2	✓	
Palula	2	✓	
Kalkoti	2	✓	
Ushojo	2	✓	
Kohistani Shina	2	✓	
Kundal Shahi	2	✓	
Shina (Gilgiti)	2	✓	
Brokskat	2	✓	
Standard Kashmiri	2	✓	

In two of the languages, Khowar and Kalasha, both belonging to the Chitral group, sex-based differentiation is entirely lacking. However, in both languages we find a two-way differentiation based on animacy, where animate nouns (including humans and higher non-human animals) are treated differently from inanimate nouns by some agreement targets. For instance, the present actual copula verb used in locational predication in Khowar has different third person

singular and plural agreement forms for animate and inanimate, respectively. That is illustrated in example (2) with the two plural forms. (The corresponding singular forms are *asúr* and *šer*.) The copula, in its various forms, is also used as an auxiliary participating in some tense-aspect formations.

(2) Khowar (Own data)

- a. *dúr-a*      *roy*      *asúni*  
 house-LOC people(AN) be.PRS.ACT.3.AN.PL  
 ‘There are people in the house.’ (KHW-PredFA:011)
- b. *kitáb*      *ma*      *dúr-a*      *šéni*  
 book(INAN) 1SG.GEN house-LOC be.PRS.ACT.3.INAN.PL  
 ‘The books are in my house.’ (KHW-PredFA:009)

A few of the dialects of NW Pashai may also lack sex-based gender distinctions (Morgenstierne 1967: 150–151); in those cases we do not have conclusive information on the presence of animacy distinctions. In another few languages – in Dameli and Shumashti (both Kunar languages), and in several of the Pashai varieties – animacy differentiation occurs, not instead of but in addition to sex-based differentiation. However, there are reasons to regard these as two separate features (with two values each) that affect different parts (or sub-domains) of the language system, a situation that Dahl (2000: 581–582) refers to as “parallel combinations of gender distinctions”. The feminine–masculine and animate–inanimate distinctions only marginally make use of the same agreement target. In Dameli, this happens in non-verbal predication, which results in a three-way differentiation at the most: animate masculine vs. animate feminine vs. inanimate, as shown in example (3). Apart from the specific domain of non-verbal predication in Dameli, a two-way masculine vs. feminine distinction is upheld in most other parts of the grammar. It is not unlikely that a similar situation holds in Shumashti, although the data available is too scanty to draw any firm conclusions.

(3) Dameli (Own data)

- a. *i*      *mač*      *mruy*      *thaa*  
 PROX.AN man(M) hunter be.PRS.3M.SG  
 ‘This man is a hunter.’ (DML-ValQuestHM:070)
- b. *poši*      *koki*      *thui*  
 cat(F) asleep be.PRS.3F.SG  
 ‘The cat is asleep.’ (DML-ErgSurvHM:013)

- c. *bum*    *šukisan daru*  
ground dry    be.PRS.3SG.INAN  
'The ground is dry.' (DML-ValQuestHM:068)

In Pashai (at least in SE, SW and NE), animacy and sex-based gender agreement do co-occur in one and the same clause and with one and the same referent, see the SE Pashai example in (12). That results in a four-way distinction (masculine/animate, masculine/inanimate, feminine/animate vs. feminine/inanimate).

This naturally leads over to the topic of our next section: agreement targets and the general pervasiveness of gender.

## 5 Agreement targets and the pervasiveness of gender

In line with the view that grammatical gender and the number of gender categories is evidenced in agreement patterns, I will use the number of agreement targets as a (somewhat crude) measure of what I call gender pervasiveness (Table 4). Here, it will be necessary to look at sex-based distinctions (masculine vs. feminine) separate from non-sex-based distinctions (animate vs. inanimate). This is not to say that they need to be regarded as two entirely distinct phenomena, but rather to underscore a general observation that sex and animacy in most cases operate at different levels and affect separate (and only peripherally overlapping) subsystems or parts of the language systems under investigation. It will be possible to make some overall generalizations along relatedness lines, although I will also point out some important variation within lower-level genealogical groupings, and for some of the languages I will also elaborate further on the relative pervasiveness within the target categories. While pronominal gender is indicated in Table 4 it will not be discussed until §7. (A tick-mark within parentheses indicates that agreement is restricted to copula verbs or copula-derived auxiliaries; a question mark after a tick-mark indicates a possible but non-conclusive presence of a gender target.)

Starting with Kashmiri, gender is very pervasive throughout the system, including adjectives, adnominal demonstratives and possessive phrases in nominal modification; verbs also show gender agreement. Person, number and gender are often conflated in a complex manner, and distinctions are, at least partly, expressed non-linearly, i.e. by vowel modification or palatalization. Example (4) demonstrates agreement in adjectival inflection; as can be seen in this example, gender distinctions are upheld in the singular as well as in the plural.



Table 4: Agreement targets for gender (sex-based, animacy-based) in Hindu Kush Indo-Aryan

Language	Gender targets									
	Sex-based					Animacy-based				
	<i>verb</i>	<i>adj</i>	<i>dem</i>	<i>poss</i>	<i>pron</i>	<i>verb</i>	<i>adj</i>	<i>dem</i>	<i>poss</i>	<i>pron</i>
Standard Kashmiri	✓	✓	✓	✓	✓					
Shina (Gilgiti)	✓	✓	✓		✓					
Brokskat	✓	✓	✓		✓					
Kundal Shahi	✓	✓								
Kohistani Shina	✓	✓			✓					
Ushojo	✓	✓			✓?					
Palula	✓	✓	✓		✓					
Kalkoti	✓	✓								
Sawi	✓	✓								
Indus Kohistani	✓	✓		✓						
Gawri (Kalami)	✓	✓		✓						✓
Torwali	✓	✓						✓?		
Bateri	✓	✓								
Tirahi	✓	✓		✓						
Wotapuri-Katarqalai	✓	✓								
Gawarbati	✓	✓		✓						
Grangali		✓								
Shumashti	✓	✓				(✓)				
Dameli	✓	✓		✓		(✓)		✓		✓
Southwest Pashai	✓	✓				(✓)				
Southeast Pashai	✓	✓				(✓)				
Northeast Pashai	✓	✓				(✓)				
Northwest Pashai	✓	✓				(✓)?				
Kalasha						(✓)				
Khowar						(✓)				

(4) Standard Kashmiri (Koul 2003: 915)

- a. *n'uul*      *kooṭh*  
blue.M.SG coat(M)  
'a blue coat'
- b. *niil*      *kooṭh*  
blue.M.PL coat(M)  
'blue coats'
- c. *niiṣ*      *kəmiiz*  
blue.F shirt(F)  
'a blue shirt'
- d. *niiṣ-i*      *kəmiiz-i*  
blue.F-PL shirt(F)-PL  
'blue shirts'

In Kashmiri, gender agreement is part of the paradigm of all major verbal categories apart from the future tense. As in Indo-Aryan in general, gender differentiation became part of the verbal paradigm as participial forms were introduced and proliferated as carriers of core tense-aspect categories during the Middle Indo-Aryan stage (Pirejko 1979: 481–482; Klaiman 1987: 61–64). In a development associated with that, the transitive subject ended up non-nominatively coded while the verb (reinterpreted as part of a finite verb construction) agreed with the nominatively coded direct object (Masica 1991: 341–346). This was the establishment of a split ergative system still in existence in various versions in many Indo-Aryan languages, including many HKIA languages (Liljegren 2014).

Gender is generally also very pervasive in the Shina group (Shina (Gilgiti) to Sawi in Table 4), although it varies between the individual languages. None of them manifest gender agreement in possessive modification. In Gilgiti Shina, Brokskat and Palula, adjectives, adnominal demonstratives and verbs are targets of gender agreement, whereas it is limited to adjectives and verbs in the rest of the languages classified as Shina. The pervasiveness of gender within the verbal paradigms varies to a great extent, and is partly related to considerable differences in verbal alignment patterns. Gilgiti Shina and Kohistani Shina, the two varieties that together constitute “Shina proper”, are characterized by consistent accusative verbal alignment in combination with ergative case marking (see example 5). A number of Shina enclaves farther to the West instead show an aspectual split between ergatively aligned clauses in the perfective (see example 6), in which the verb agrees in gender and number with the direct object, and

accusatively aligned clauses in the non-perfective. In Shina proper, gender agreement is largely conflated with person-marking, whereas in the Western varieties, gender- and number-inflected verb forms (based on participles) have largely replaced person-inflected forms.

- (5) Gilgiti Shina (Own data)

*ro baál-se khirkí phuṭ-eég-u*  
 REM.M.SG boy(M)-ERG window(F) break-PFV-3M.SG  
 ‘The boy broke the window.’ (SCL-ValQuestAH:025)

- (6) Palula (Own data)

*phoo-á darúri phooṭéel-i*  
 boy(M)-OBL window(F) break.PFV-F  
 ‘The boy broke the window.’ (PHL-ValQuestNH:025)

In addition to the categories surveyed in this section, gender agreement in Palula is also extended or copied to e.g. adjuncts in predicatively used adverbial phrases. In (7), the scalar modifier *búḍ*- ‘much’ agrees with the feminine noun head of the subject.

- (7) Palula (Own data)

*asíi iskuúl bi asaám the búḍ-i dhúura hín-i*  
 1PL.GEN school(F) also 1PL.ACC to much-F distant be.PRS-F  
 ‘Our school is also very far away for us.’ (PHL-OUR:016)

In none of the Kohistani languages are adnominal demonstratives targets of gender marking. On the other hand, gender differentiation is part of possessive modification in at least two of the languages. Examples are provided from Indus Kohistani in (8).

- (8) Indus Kohistani (Lubberger 2014: 62, 82)

- a. *zâĩ bakàr*  
 1PL.POSS.F goat(F)  
 ‘our goat’  
 b. *zâã baá*  
 1PL.POSS.M house(M)  
 ‘our house’

Manifestation of gender in the verbal paradigm is not necessarily much less pervasive than in the languages of the Shina group, but it tends to be more challenging in terms of description. It is to a greater extent non-segmental in Kohistani than in Shina. A case in point is the Kohistani language Gawri (a.k.a. Kalam Kohistani) which historically has lost most of its gender-specific endings (both on the nouns themselves and on their agreement targets) as well as its suffixing plural or case-marking. It has, however, preserved the distinctions themselves up to a point, in the form of vowel modifications and/or distinct tonal patterns, as can be seen in example (9).

(9) Gawri

- a. Inflection of nouns (H=high tone, LH=low to high, HL=high to low, L=low) (Baart 1999: 36)

SG.NOM		PL.NOM/SG.OBL/PL.OBL		
<i>šaak</i>	H	<i>šääk</i>	HL	‘piece of wood’ (M)
<i>dätär</i>	LH	<i>dätär</i>	L	‘cooking frame’ (M)
<i>naar</i>	H	<i>neer</i>	HL	‘root’ (F)
<i>däriin</i>	LH	<i>däriin</i>	L	‘ground’ (F)

- b. Gender and number agreement on adjectives (Baart 1999: 19; p.c. Muhammad Zaman Sagar)

<i>raan</i>	<i>poo</i>	<i>rään</i>	<i>lukuṭor</i>
good.M.SG	boy	good.M.PL	boy.PL
‘good boy’		‘good boys/children’	
<i>reen</i>	<i>bire</i>	<i>reen</i>	<i>likiṭeer</i>
good.F	girl	good.F	girl.PL
‘good girl’		‘good girls’	

- c. Gender and number agreement on verbs (conflated with aspect marking) (Baart 1999: 19; p.c. Muhammad Zaman Sagar)

<i>poo</i>	<i>bäch-an-t</i>	<i>lukuṭor</i>	<i>bäch-än-t</i>
boy	go-IPFV.M.SG-PRS	boy.PL	go-IPFV.M.PL-PRS
‘The boy is going.’		‘The boys are going.’	
<i>bire</i>	<i>bäch-en-t</i>	<i>likiṭeer</i>	<i>bäch-en-t</i>
girl	go-IPFV.F-PRS	girl.PL	go-IPFV.F-PRS
‘The girl is going.’		‘The girls are going.’	

Masculine and feminine agreement forms are clearly distinguished in all of the major tense-aspect categories in Gawri and Torwali, either inflectionally or by vowel alternation. However, a high degree of levelling seems to have taken place in Indus Kohistani; and most likely in Bateri too. In Indus Kohistani and Bateri, transitive verbs (or at least most of them) are invariant in the simple past (i.e., there is no agreement with any of the arguments). In addition, the application of the ergative marking of the transitive subject is variable. In Bateri, a nominative vs. ergative contrast is possibly missing altogether with full nouns, as evidenced in example (10).

(10) Bateri (Own data)

- a. *yak muuʃ as-uu*  
one man(M) be.PST-M.SG  
'There was a man.' (BTV-PearStoryMB:001)
- b. *muuʃ ɖaʌn sand-id*  
man(M) stick make-PST  
'The man made a stick.' (BTV-ValQuestMB:085)

In the Kunar group, the targets of sex-based gender differentiation are adjectives, verbs and, in the case of Gawarbatl and Dameli, possessive modifiers. The sentences in (11) illustrate some of those agreement patterns in Gawarbatl: possessive and verbal (copula) agreement with a feminine noun in (a), possessive agreement with a masculine noun in (b), and adjectival and verbal agreement with a feminine noun in (c). Verbal agreement that takes gender into account is rather restricted in Gawarbatl: it occurs only with intransitive verbs, and for third person singular. As seen in (b), the transitive subject in the past (perfective) is ergatively marked, while verbal agreement is accusatively aligned.

(11) Gawarbatl (Own data)

- a. *woi ʃekura-an-i awaaz then-i*  
PROX.SG boy(M)-POSS-F voice(F) be.PRS-3F.SG  
'This is a boy's voice.' (GWT-NPhonNU:071-4)
- b. *ʃekuri-e kitaab-an-a faʃaa daal-us*  
girl-ERG book(M)-POSS-M leaf(M) tear-PST.3SG  
'The girl tore the page from the book (lit. the book's leaf).'  
(GWT-ValQuestAS:032)

- c. *pol-i    tekuri hans-ui*  
 small-F girl(F) laugh-PRS.3F.SG  
 ‘The little girl laughed.’ (GWT-ValQuestAS:057)

As already mentioned in §4, an added distinction between animate and inanimate occurs in Dameli and Shumashti. While animacy influences lexical or constructional choices on various levels of Dameli, the only purely paradigmatic contrasts that depend on animacy values are those of the copula verb (Perder 2013: 121–125), as illustrated above in example (3), and of demonstratives. However, it is highly uncertain whether the inanimate copula is at all used as an auxiliary in verbal predication in any of the tense-aspect categories in Dameli. More interestingly, Perder (2013: 51–55) observes what seems to be an ongoing restructuring of the entire gender system, a point to which we shall return in the next section when discussing assignment criteria.

In Pashai, sex-based gender is again relatively pervasive, although limited in its manifestation to adjectives and verbal agreement. As in Dameli, there is an additional layer of animacy-based differentiation in the verbal paradigm. Lehr (2014: 255) describes (for SE Pashai) how the masculine vs. feminine distinction is upheld throughout the past and perfective parts of the verbal paradigm, a contrast that is present in first, second as well as in third person. The additional animate vs. inanimate distinction, on the other hand, is limited to the verbal system (2014: 256–257), occurring only in non-verbal predication and in the (participial-based) present perfect category. The three sentences in (12) are all examples of the present perfect: the main verb agrees in person with the subject, in sex-based gender with the object, and the auxiliary agrees in sex-based as well as non-sex-based gender and person with the object.

(12) SE Pashai (Lehr 2014: 290, 297)

- a. *pari-y      kelaa    kaṭ-ee=šeer-a      ne-l-aw-aa-e*  
 Pari(F)-OBL boy(M) cot-OBL=head-LOC sit-TRZ-STV.PTC-M-POSS.3SG  
**aas**  
 be.AN.M.PRS.3  
 ‘Pari has seated the boy on the cot.’
- b. *miy            maada-y      doa      be    ka-w-aa-e*  
 DEM.SG.OBL woman(F)-OBL prayer(M) too do-STV.PTC-M-POSS.3SG  
*š-i*  
 be.INAN.PRS-3  
 ‘This woman has made a prayer.’

- c. *mam pelek meez-ee=šeer-a      ĵe-w-i-m*  
 I      cup(F) table(F)-OBL=on-LOC place-STV.PTC-F-POSS.1SG  
*š-i*  
 be.INAN.PRS-3  
 ‘I have placed the cup on the table.’

Finally, both of the two Chitral group languages, Khowar and Kalasha, entirely lack any sex-based gender in their agreement patterns. Grammatical differentiation between animate and inanimate nouns is manifested, but only in the verbal paradigm. It occurs in those verbal categories that are constructed with a copula-based auxiliary, such as in the Kalasha example in (13): here, the animate as well as the inanimate forms occur, each along with the main verb ‘hit’. Kalasha expresses animate vs. inanimate differentiation in five of its nine main tense-aspect categories (Bashir 1988: 60–72), but because of its consistent accusative alignment with subject agreement (as compared to the pattern of direct object agreement in Pashai), the frequency of inanimate marking is in effect rather low. A similar situation holds for Khowar (Bashir 1988: 123–133). Thus, the centrality of the animacy contrasts that these tense-aspect systems allow for could in fact be questioned.

- (13) Kalasha (Heegård Petersen 2015: 250)  
*gheri tya-y      a-aw=e,      tasa      ek bab-as*  
 again hit-PFV.PTC AUX.AN.ACT-3SG=when 3SG.REM.OBL a sister-OBL.SG  
*gulin-a tya-y      š-iu.*  
 lap-LOC hit-PFV.PTC AUX.INAN-PRS/FUT.3SG  
 ‘When he hit [the ball] again, it was hit into her sister’s lap.’

It seems that whereas sex-based gender generally is deeply entrenched in the languages that have it, and is clearly evidenced in many of the inflectional paradigms, the non-sex based type of gender differentiation that we saw examples of in a few of the languages is indexed in considerably fewer domains and is thus affecting, in each case, a rather limited domain of the language system. The question remains open as to whether those contrasts should be seen as instances of mere (lexical) co-occurrence restrictions, instead of truly grammatical contrasts. We may also regard the occurrence of animacy distinctions in these languages as examples of overdifferentiated targets (Corbett 1991: 168–169), probably more so in the languages with parallel combination of distinctions (Dameli, Shumashti and the Pashai varieties) than in the languages with non-sex based distinctions only (Khowar and Kalasha).

## 6 Assignment criteria

Determining the assignment criteria for gender in individual languages is a less straightforward matter, even for much more well-known languages with large corpora available. For this reason, the following is meant only as a very tentative assessment, and the results of the assessment is therefore not reduced to a simple table representation. Although the focus will be on the languages for which there is a more comprehensive description in place, it remains beyond the present investigation to lay down precise assignment rules for any of these.

For all the languages that have a sex-based two-term system, i.e. the large majority of HKIA, gender is with high consistency assigned according to natural sex as far as nouns denoting humans and other higher animates, particularly domestic animals, are concerned. Below this cut-off point between higher and lower animates (or possibly between animates and inanimates), semantics is a much less reliable indicator, although some outstanding semantic properties beside sex will be mentioned in connection with the discussion of individual languages. But it also seems clear that formal (i.e. non-semantic) criteria do play a non-trivial role in some of the languages in assigning inanimate and lower animate nouns to the masculine and feminine classes, respectively. In a historical perspective, the present two-term systems is the result of the masculine and the neuter categories of the former three-gender system having merged (Masica 1991: 221). This, however, is not mirrored in a totally unbalanced feminine to masculine ratio, as might be expected. Instead, there is a relatively even distribution; in Palula, there were 58 per cent masculine and 42 per cent feminine nouns in a database comprising about 1,300 nouns, and in a Gawri list of 2,000 nouns, the percentages were 60 and 40, respectively (Baart 1999: 82), and inanimates and lower animates of both genders are numerous.

Although there are plenty of examples in Kashmiri of feminine nouns derived from masculine nouns by means of various semi-regular phonological processes (such as stem vowel diphthongization or fronting) these correlations between characteristic phonological features and one or the other gender are mainly restricted to higher animates: *guur* ‘milkman’ vs. *guuər* ‘milkwoman’; *koɬ* ‘boy’ vs. *kəɬ* ‘girl’; *kəkur* ‘rooster’ vs. *kəkir* ‘hen’; *mool* ‘father’ vs. *məj* ‘mother’. However, the nominal inflectional patterns of the language (see Table 5) also predict gender to a large extent. Most non-nominative case forms, for instance, have endings that are typical for masculine vis-à-vis feminine nouns (with a great deal of syncretic *i* occurring in the paradigms of feminine nouns, contrasting with differentiating forms in the paradigms of masculine nouns), often accompanied by stem alternations (with vowel fronting or palatalization in the feminine forms).



Table 5: Sample Kashmiri nominal paradigm (Koul 2003: 909)

Case	'boy' M		'girl' F	
	SG	PL	SG	PL
NOM	<i>lɔḍki</i>	<i>lɔḍki</i>	<i>kuur</i>	<i>koori</i>
DAT	<i>lɔḍkas</i>	<i>lɔḍkan</i>	<i>koori</i>	<i>kooren</i>
ERG	<i>lɔḍkan</i>	<i>lɔḍkav</i>	<i>koori</i>	<i>koorev</i>
GEN	<i>lɔḍki</i>	<i>lɔḍkan</i>	<i>koori</i>	<i>kooren</i>

In the Shina group, many of the languages have sizeable subclasses of masculine and feminine nouns with gender-typical endings, mostly *o/u/a* with masculine nouns, and *i* with feminine nouns. But again, similar to what was noted regarding Kashmiri, there is a considerable overlap between nouns with such overt gender markers and biological sex. Brokskat, a Shina language which otherwise has few overt phonological characteristics related to one or the other gender, makes use of two Tibetan-derived suffixes, *-pa/-po* and *-ma/-mo* to indicate the sex of some higher animates (see Table 6). To what extent these suffixes are used with inherited vocabulary is not clear.

Table 6: Masculine–feminine higher animate pairs in Brokskat (Ramaswami 1982: 38–39; Sharma 1998: 56–58, 80)

Masculine		Feminine	
<i>rgəl-po</i>	'king'	<i>rgəl-mo</i>	'queen'
<i>bäg-pa</i>	'bridegroom'	<i>bäg-ma</i>	'bride'
<i>bya-po</i>	'rooster'	<i>bya-mo</i>	'hen'
<i>əbs</i>	'horse'	<i>əspi, rgun-ma</i>	'mare'
<i>byo</i>	'boy, son'	<i>mole</i>	'girl, daughter'
<i>dudo</i>	'grandfather'	<i>dede</i>	'grandmother'
<i>čhatalo</i>	'he-goat'	<i>aav</i>	'she-goat'
<i>laanto</i>	'bull'	<i>gooli</i>	'cow'

However, for many consonant-ending nouns below the threshold for sex-based assignment, i.e. between higher and lower animates, assignment seems to a large extent arbitrary in Shina languages. Although there are clearly discernible declensional classes in e.g. Kohistani Shina, Palula and Sawi, these are not in all cases directly mapped to one or the other gender. In Gilgiti Shina, a language

where declensional differences are less clearly identifiable, there are fewer formal clues to gender assignment, and in Brokskat, where there are few phonological clues and a relatively uniform inflectional pattern, the arbitrariness seems even more noticeable as far as nouns low on the animacy scale are concerned. It is in fact likely that gender assignment in these languages to a varying extent is an intricate interplay of overlapping semantic, morphological and phonological factors, not altogether different from what we find in e.g. German (Corbett 1991: 49).

Let us take Palula as an example in terms of such a complex interplay of different assignment criteria. Starting with nominal morphology (see Table 7), Palula has three major declensional classes, characterized by plural formation with *-a*, *-i* and *-m*, respectively. The *m*-declension consists exclusively of feminine nouns (all of which end with gender-typical *i* in their singular form), whereas *a*-declension consist to 79 per cent of masculine nouns, and the *i*-declension to 70 per cent of feminine nouns. In addition, there are two minor declensions (together representing 10–15 per cent of all nouns), both exclusively masculine.

Table 7: Palula noun declensions

Decl	SG NOM	SG OBL	PL NOM	PL OBL	Relative size	M/F
<i>a</i> -decl	<i>púustu</i> 'skin (M)'	<i>púust-a</i>	<i>púust-a</i>	<i>púust-am</i>	large (~50%)	79/21
<i>i</i> -decl	<i>baát</i> 'word (F)'	<i>beet-í</i>	<i>beet-í</i>	<i>beet-íim</i>	large (~25%)	30/70
<i>m</i> -decl	<i>ťiki</i> 'bread (F)'	<i>ťiki</i>	<i>ťiki-m</i>	<i>ťiki-m</i>	large (~13%)	0/100
<i>ee</i> -decl	<i>alučá</i> 'plum (M)'	<i>alučá</i>	<i>aluč-eé</i>	<i>aluč-eém</i>	small (~8%)	100/0
<i>aan</i> -decl	<i>đaakú</i> 'robber (M)'	<i>đaaku-á</i>	<i>đaaku-aán</i>	<i>đaaku-aanóom</i>	small (~5%)	100/0

However, the amount of arbitrariness within the two “gender-divided” declensions is further reduced by taking phonological clues into account (see Table 8). About a third of the nouns in the *a*-declension have for Palula gender-typical endings in their nominative singular forms (mainly masculine nouns in *u*, and feminine nouns in *ái*). A typical property of many *i*-declension consonant-ending nouns that are assigned feminine gender is that they have a second-mora accented *aá* which very often is subject to a process of umlaut (> *ee*) in its inflected

Table 8: Gender-typical phonological properties in Palula

Masculine			Feminine		
Cu#	<i>tómbu</i> ‘trunk’, <i>šúuru</i> ‘hole’, <i>rúulu</i> ‘tear’, <i>púustu</i> ‘skin’, <i>príiṣu</i> ‘flea’, <i>báabu</i> ‘father’, <i>báatru</i> ‘irrigation lock’, <i>bháaru</i> ‘load’	<i>a</i> -decl	Ci#	<i>šúri</i> ‘ladder’, <i>ṭíki</i> ‘bread’, <i>šiśáki</i> ‘ogress’, <i>phéepi</i> ‘father’s sister’, <i>nóki</i> ‘beak’, <i>múṭi</i> ‘arm’, <i>lúuṭi</i> ‘ball of yarn’, <i>bééji</i> ‘heifer’	<i>m</i> - decl
Coó#	<i>rhoó</i> ‘song’, <i>phoó</i> ‘boy’, <i>paṇoó</i> ‘slipper’, <i>muuṣoó</i> ‘elbow’, <i>baḍiloó</i> ‘male descendant of Badil’, <i>haṇoó</i> ‘egg’	<i>ee</i> - decl, <i>a</i> -decl	Cíi#	<i>rhootašii</i> ‘morning’, <i>rhaíi</i> ‘footprint’, <i>phaaṭurii</i> ‘butterfly’, <i>aḥíi</i> ‘eye’, <i>balíi</i> ‘roof end’, <i>bíi</i> ‘seed’	<i>a</i> -decl
Cá#	<i>ṭeeká</i> ‘contract’, <i>lambá</i> ‘flame’, <i>alaaqá</i> ‘area’, <i>alučá</i> ‘plum’, <i>çaṇžá</i> ‘torch’	<i>ee</i> - decl, <i>i</i> -decl	Cái#	<i>tookrái</i> ‘basket’, <i>puṭái</i> ‘piece of meat’, <i>mulái</i> ‘radish’, <i>bhraajái</i> ‘sister-in-law’	<i>a</i> -decl
Caá#	<i>saaraá</i> ‘wilderness’, <i>raajaá</i> ‘ruler’, <i>paalaá</i> ‘leaf’, <i>aaghaá</i> ‘sky’, <i>bhalaá</i> ‘evil spirit’, <i>čoolaá</i> ‘speech, style’	<i>i</i> -decl, <i>ee</i> - decl	CaáC#	<i>aašaár</i> ‘apricot’, <i>salaám</i> (pl. <i>saleemí</i> ) ‘greeting’, <i>oombaár</i> (pl. <i>oombeerí</i> ) ‘canal inlet’, <i>baát</i> (obl. <i>beetí</i> ) ‘word’	<i>i</i> -decl

forms (with affixes involving *i*). This is also characteristic of a good number of loan words. This is not to say that there are no exceptions to these correlations between certain vocalic properties and one of the two genders, but they are indeed few.

Another sizeable group of *a*- and *i*-declension nouns (although partly overlapping with those having gender-typical phonological properties) are assigned

gender semantically. Primarily that is by biological sex for nouns referring to humans and higher non-human animates. Word pairs referring to male and female, respectively, which have a common lexical root are frequent (see Table 9), especially in the realm of kinship. For most higher animates, the masculine is the default, and for those that have a feminine counterpart, the latter is a marked form (often part of the *m*-declension and ending in *i*), i.e. the one used only when a specification of sex is called for. However, in a few cases, the reverse holds, e.g. with ‘fox’ and ‘cat’. The semantic relationship between masculine ‘goat kid’ and its feminine counterpart ‘goat (generic)’ is again different.

Table 9: Masculine–feminine higher animate pairs in Palula

Masculine		Feminine	
<i>jáanu</i>	person	<i>jëeni</i>	female person
<i>saaróonu</i>	woman’s sister’s husband	<i>saaréeni</i>	wife’s sister
<i>phoó</i>	boy	<i>phaí</i>	girl
<i>móomu</i>	mother’s father	<i>méemi</i>	mother’s mother
<i>káaku</i>	older brother	<i>kéeki</i>	older sister
<i>khaamaád</i>	owner, husband	<i>khaaméedi</i>	female owner
<i>práachu</i>	guest	<i>préeči</i>	female guest
<i>phóopu</i>	father’s sister’s husband	<i>phéepi</i>	father’s sister
<i>kučúru</i>	dog	<i>kučúri</i>	female dog
<i>bačhúuru</i>	young calf	<i>bačhúuri</i>	young female calf
<i>karáaru</i>	leopard	<i>karéeri</i>	female leopard
<i>iñç</i>	bear	<i>iñçi</i>	she-bear
<i>luumóo</i>	male fox	<i>luumái</i>	fox (generic)
<i>púšu</i>	tom-cat	<i>púši</i>	cat (generic)
<i>kakóok</i>	chicken	<i>kakuéeki</i>	hen
<i>čhaál</i>	goat kid	<i>čhéeli</i>	goat (generic)

Apart from this relatively straightforward correlation between sex and grammatical gender, there is another (but obviously related) correlation, namely between relative size or power and gender, primarily applied to lower animates and inanimates (as exemplified in Table 10). In these cases, the derivation of feminine nouns could be described as a type of diminutive formation. The similarity in kind

is more approximate and less predictable than with the previously exemplified higher animate pairs.

Table 10: Masculine–feminine lower animate and inanimate pairs in Palula

Masculine		Feminine	
<i>phútu</i>	fly	<i>phúti</i>	mosquito
<i>khaláaru</i>	large leather bag, made from skin of a he-goat	<i>khaléeri</i>	small leather bag, made from skin of a she-goat
<i>şúuru</i>	hole	<i>şúuri</i>	cap
<i>angúru</i>	thumb, big toe	<i>angúri</i>	finger, toe
<i>açhibáaru</i>	eyebrow	<i>açhibéeri</i>	eyelashes
<i>angóor</i>	fire	<i>angeerí</i>	charcoal

Leaving Palula and the Shina languages for now, some of the languages of the Kohistani group also have overt phonological markers, similar to the ones in the Shina group. In Indus Kohistani, *i*-endings are associated with a group of feminine nouns, and in Bateri some masculine nouns end in *-o/-u* and some feminine nouns in *-a/-ã*. In both of these cases, however, that pattern is relatively restricted and perhaps primarily relevant for feminine nouns derived from masculine nouns denoting humans, particularly applied to male–female pairings in the kinship systems of these languages. Due to historical loss of final vowel segments, the corresponding correlations in Gawri and Torwali are often only preserved in stem vowel alternations and tonal contrasts, resulting from assimilation prior to apocope. In Gawri, there is a strong correlation between feminine gender and the vowel qualities [i] and [e], and a corresponding correlation between masculine gender and the qualities [a], [æ], [o], and [u].

In the Kunar languages, there are no obvious declensional differences (plurality is for instance normally left morphologically unmarked, and case marking has little allomorphy), and nouns that have gender-typical endings are relatively few (*a*-ending masculine nouns in Dameli, Gawarbati and Shumashti; *i*-ending feminine nouns in Dameli and Gawarbati; *i*-ending or *ik*-ending feminine nouns in Shumashti). Like in many of the other groups, nouns with these overt phonological “markers” often participate in masculine–feminine pairings where the latter term is derived from the former, which frequently applies to humans or domestic animals. Although needing a more systematic study, there is

evidence suggesting that Dameli is drifting away from formal-semantic gender assignment toward purely semantic gender assignment, as strict masculine vs. feminine gender assignment is becoming restricted to nouns above the cut-off point between higher and lower animates. This is for instance manifested in the native speaker inconsistency that Perder noted while eliciting the gender of inanimate nouns (2013: 54), along with an observed pattern of a default application of masculine gender agreement between verbs and inanimate subjects (2013: 111). Together with the already-mentioned observations regarding animacy-related distinctions, it seems like we are witnessing a development in Dameli from a partly formal assignment system with two sex-based grammatical genders to a system by which gender is assigned entirely along semantic lines. In most parts of the system there is a contrast between a feminine class consisting of female higher animate nouns and a masculine class with all the remaining nouns, and in a restricted part of the system (with the copula verb as target) there is a three-way contrast between higher animate males, higher animate females and the rest. The grammatical animate-inanimate distinction in Dameli is, as far as has been observed, altogether missing in Gawarbatī, leaving it with a two-way distinction and with assignment principles along the same lines as described for many of the Kohistani and Shina languages. Although the scanty material available does not give us any firm evidence, the Shumashti copula forms that Morgenstierne (1945: 255) presents us with (*in-e* ‘is M’, *in-i* ‘is F’, *šuu-e* ‘(it) is’) implies an actual four-way differentiation, although we can only assume that a hypothetical inanimate feminine form (\**šuu-i* ‘(it) is F’) simply is missing in the data.

The patterns observed for most parts of the other groupings can also be seen in Pashai. Here, too, there are certain endings associated with one or the other gender. In SE Pashai, for instance, *-i* or *-ek* is typical of feminine nouns and *-aa* of masculine. While the feminine *i*-ending is found with many inanimate nouns, there are many regular alternations involving gendered pairs where the masculine form with *-aa* contrasts with a feminine form with *-ek*. But again, there are numerous nouns that are either masculine or feminine that have none of these overt phonological markers. Nor is there much in terms of declensional differences. The only clear distinction in plural marking is instead related to humanness or animacy. The choice of copula and auxiliary forms is, like in Dameli, entirely governed by semantics. This gives us in effect a system of two sex-based genders, masculine and feminine, each with two sub-genders, animate and inanimate.

The assignment in the languages of the Chitral group, which are entirely void of any sex-based distinctions, goes only along semantic lines, where the auxiliary

use in the verbal paradigms reflects an animate vs. inanimate distinction. Certain local case markers only occur with inanimate nouns and not with animate nouns (Heegård Petersen 2006: 53; Bashir 2003: 844). However, it is doubtful whether this can be considered a primary assignment criterion.

## 7 Pronominal gender

A separate issue, but also necessary to mention in the context, is the presence of pronominal gender distinctions in Hindu Kush Indo-Aryan. In pronominal gender (see Table 4) we find some interesting differences, partly going along sub-classification lines. Even in this case, it is more instructive to differentiate between sex-based distinctions and non-sex-based (i.e. animacy-based) distinctions. Interestingly, so far, no combination of the two (in the same domain) has been noted for any individual language. Note, that only personal pronouns (or demonstratives used as third person pronouns) have been taken as diagnostic in this case.

Only in two of the subgroups do we find evidence for differentiating personal pronouns for masculine and feminine referents (including non-human animates and inanimates), in Kashmiri and in at least four of the Shina languages. These languages all have a two-term system, a masculine third person pronoun contrasting with a feminine, so that even reference to inanimates makes use of one of the two according to their grammatical gender. The differentiation is limited to singular reference and third person, whereas the same term is used for masculine plural and feminine plural alike. Gender is also neutralized in some of the case forms. For instance, Kohistani Shina (14), has separate feminine (a) and masculine (b) ergative pronouns for perfective transitive constructions, whereas there is only one third person singular form used in non-perfective transitive constructions (c) or in intransitive clauses (d).

(14) Kohistani Shina (Schmidt & Kohistani 2008: 181, 217, 247, 224)

- a. *séso*                      *asór*    *ṭiki*    *d-eég-i*.  
3F.SG.ERG.PFV 1PL.DAT bread give-PFV-3F.SG  
'She gave us food.'
- b. *sési*                      *ráaty-oo*    *kom*    *th-áa-o*.  
3M.SG.ERG.PFV night-ABL work do-PFV-3M.SG  
'He worked all night.'

- c. *ses*            *dōóchi*    *ágo*            *çiç-eé*            *táam*  
 3SG.ERG.IPFV tomorrow headshawl embroider-CV complete  
*th-úu.*  
 do-FUT.3F.SG  
 ‘She will finish embroidering the headshawl tomorrow.’
- d. *sa*            *ruleé*    *b-eé*    *boǰ-áa-n-i.*  
 3SG.NOM disguise be-CV go-IPFV-AUX.PRS-3F.SG  
 ‘She goes (there) disguised.’

Within the Shina group, there are four different patterns (see Table 11). In Gilgiti Shina and in Brokskat, both nominative and ergative have distinct masculine and feminine forms. In Kohistani Shina (as illustrated above), this distinction is upheld in the (perfective) ergative but is neutralised in the nominative (and elsewhere). In Palula, the opposite holds, and it is in the nominative that gender is differentiated whereas it is neutralised in the ergative (and elsewhere). In Sawi, Kalkoti, Kundal Shahi and possibly in Ushojo, no pronominal gender differentiation is made at all. Kashmiri, the only other HKIA language that makes pronominal gender distinctions, displays the same pattern as Gilgiti Shina does.

Table 11: Pronominal third person gender distinctions in Shina languages

	Nominative		Ergative	
	Masc.	Fem.	Masc.	Fem.
Gilgiti Shina	<i>ro</i>	<i>re</i>	<i>ros</i>	<i>res</i>
Kohistani Shina	<i>sa</i>		<i>sési</i>	<i>séso</i>
Palula	<i>so</i>	<i>se</i>		<i>tíi</i>
Sawi		<i>see</i>		<i>ti</i>

Pronominal differentiation related to animacy is found in a few individual languages belonging to different subgroups. Different pronouns for animate and inanimate reference, respectively, are used in Gawri, as in example (15), in Dameli and possibly also in Torwali.

(15) Gawri (Baart & Sagar 2004: 35, 52)

- a. *ääs*            *sä*    *äsēē*            *duu isaal*    *yeeš.*  
 3SG.OBL.VIS.AN with 3SG.VIS.POSS.F two women come.PFV.F.PST  
 ‘Both his wives had also come with him.’



- b. *abdul häq-ẽẽ      ǎn                      mǎy yǎrǎáz nǎāt*  
 Abdul Haq-POSS.F 3SG.OBL.VIS.INAN in interest is.not  
 ‘For Abdul Haq, there is no interest in it.’

Curiously, such a distinction is not found in the two languages that otherwise make the most systematic use of animacy distinctions in their agreement patterns, Kalasha and Khowar. For the latter, see example (16).

(16) Khowar (Own data)

- a. *awá      ho                      mar-ít-am*  
 1SG.NOM 3SG.DIST.OBL kill-PST.ACT-1SG  
 ‘I killed him.’ (KHW-PronDemAA:010)
- b. *tu      ho                      paš-ís-an-a*  
 2SG.NOM 3SG.DIST.OBL see-2SG-PRS/FUT.SPC-Q  
 ‘Can you see that? [the speaker pointing to an object a few feet away]’ (PronDemAA:018)

## 8 Gender complexity

Based on the findings in §4–§7, a cautious attempt is made at measuring the relative complexity of the gender systems in HKIA, guided by the complexity metric as laid out by Di Garbo (2016), based on the three following dimensions of complexity: the number of values, the number and nature of assignment rules, and the amount of formal marking, as previously proposed by Audring (2014). In order to arrive at a more significant internal differentiation between the HKIA languages than would otherwise be the case, the metrics were slightly adjusted (see Table 12) as compared to Di Garbo’s. Di Garbo’s features related to manipulable assignment and cumulative exponence, were for instance not taken into account here, partly due to non-applicability to the languages of my sample, partly due to unavailability of comparative data. In the case of the values dimension, a language with four or more genders receives the maximum score (instead of those with 5 or more), and in the case of indexation domains, a language with five or more targets receives the maximum score (instead of those with 4 or more). It is therefore important to note that the scores are primarily intended to provide a relative (i.e. sample-internal) measure (min=0, max=1) rather than being comparable in a wider cross-linguistic sense.

complexity languages.

This metric has been applied to each of the HKIA languages, resulting in the ranking displayed in Table 13. For some of the languages, the number of genders

Table 12: Gender complexity metric (as applied to HKIA)

Complexity dimension	Values	Score
Number of genders	Two genders	0
	Three	0.5
	Four or more	1
Number/nature of assignment rules	Semantic or formal	0
	Semantic + formal	1
Number of target domains	One target domain	0
	Two	0.25
	Three	0.5
	Four	0.75
	Five or more	1

(see Table 3) varies between dialects or is not entirely clear from the descriptions available. In those cases, the highest number in a range was used in the calculation. As for the number of target domains (see Table 4), no differentiation was made between sex-based and non-sex-based agreement. To counter a too literal interpretation of the individual complexity scores, the languages have been grouped into three complexity categories: those scoring *up to and including* 1/3 are LOW gender complexity, those scoring *more than* 1/3 *up to and including* 2/3 are MEDIUM, and those scoring *more than* 2/3 are HIGH gender

In the high complexity category we find three of the four Pashai languages and Shumashti, i.e. the only languages in our sample where we may (although far from conclusively) speak of four genders, or rather systems in which animacy and sex-based differentiation overlap; and Kashmiri, the latter a two-gender system characterized by a high number of target domains. At the other extreme, that is the low complexity category, we find Khowar and Kalasha, the only two languages in our sample with a purely semantic two-way (animate-inanimate) differentiation, as well as Grangali, a masculine-feminine-gender language characterized by having only a single agreement domain. The remaining 17 languages are all of medium complexity according to this metric.

However, it is important to point out that there are other (less measurable) factors, not included in the present metric, that contribute to the overall complexity of individual gender systems, such as the interplay between different assignment criteria (briefly mentioned in §6), declensional differences that do not map directly onto gender distinctions, and the conflation of gender and other grammatical categories (e.g. number and case).

Table 13: HKIA languages ranked for complexity

Rank	Language	Complexity score	Complexity category
1	SW Pashai	0.75	High
1	SE Pashai	0.75	
1	NE Pashai	0.75	
1	Shumashti	0.75	
2	Kashmiri	0.67	
3	Gawri	0.58	Medium
3	Indus Kohistani	0.58	
3	Brokskat	0.58	
3	Palula	0.58	
3	Shina (Gilgiti)	0.58	
4	Tirahi	0.50	
4	Torwali	0.50	
4	Dameli	0.50	
4	Gawarbati	0.50	
4	Ushojo	0.50	
4	Kohistani Shina	0.50	
5	NW Pashai	0.42	
5	Bateri	0.42	
5	Wotapuri-Katarqalai	0.42	
5	Kalkoti	0.42	
5	Kundal Shahi	0.42	
5	Sawi	0.42	
6	Grangali	0.33	Low
7	Khowar	0.00	
7	Kalasha	0.00	

## 9 Distribution and areal-linguistic implications

The findings presented above enable us to present at least some general tendencies in the geographical distribution of gender properties (see Figure 2).

First, a sex-based gender system with the two values masculine and feminine is the default for Hindu Kush Indo-Aryan. Such a system is found throughout the region, from east to west. However, two exceptions were noted, Khowar and Kalasha, where sex-based differentiation is lacking altogether. Both are situated at the northwestern periphery of the Hindu Kush region, representing the ultimate frontier of Indo-Aryan in general. Furthermore, it is in an adjacent area to those two languages that we find Dameli, a language where sex-based gender is described as being on the retreat. In at least some dialects of NW Pashai, another language spoken in the western-most part of Hindu Kush, sex-based gender may be altogether absent. Non-sex-based gender, or more specifically gender distinctions that have a contrast between animate and inanimate at their core, are also represented in the region, but only clearly so in the western part of the region. Two of the languages with such a basis are, again, Khowar and Kalasha. In a few other languages spoken in the vicinity of the former two – most prominently in varieties of Pashai – an animacy-based system overlaps with a sex-based system. However, the targets for such gender distinctions are often kept distinct.

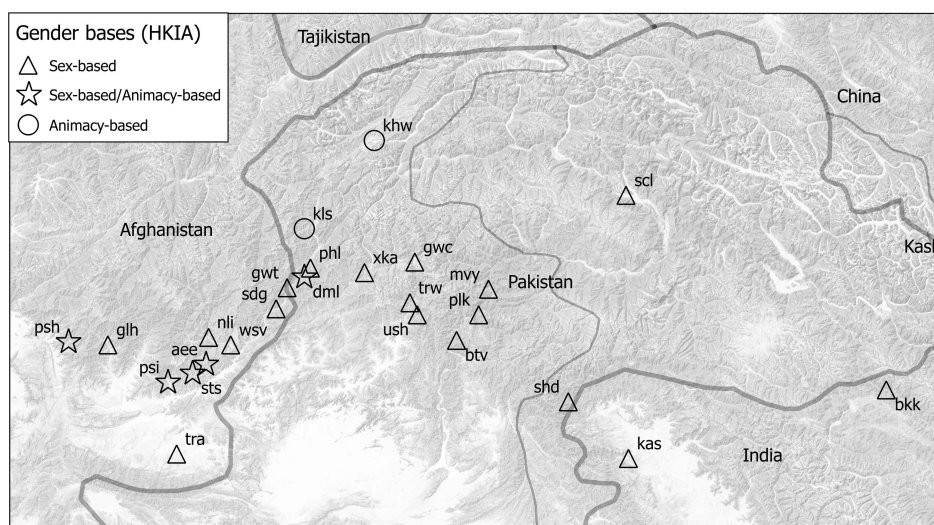


Figure 2: Gender bases in HKIA languages

Second, gender is generally deeply entrenched in those languages that have a sex-based system. Especially in Kashmiri and Shina, i.e. the languages mainly

spoken in the eastern part of the region, gender agreement is displayed with a wide range of targets. In a number of those languages, it is intertwined with person agreement in their verbal morphology, and we also noted some examples of gender agreement being extended to further targets. Kashmiri and some of the Shina languages have gender agreement with demonstratives, and it is only in these languages that we also find sex-based pronominal gender. Gender in some of the Kohistani languages, spoken in the central part of the region, is almost equally pervasive. However, the lack (or loss) of direct object agreement in a few of those languages and the subsequently lower frequency of gender agreement with noun phrases low in the animacy hierarchy may in the long run weaken the masculine–feminine differentiation in parts of the vocabulary where sex plays no role in assignment. Accusative verbal alignment, along with relatively few agreement targets, is probably in some ways related to the erosion of sex-based gender in the Kunar languages in the western part of the region.

In Kashmiri, Kohistani and Kunar, possessive modifiers are frequently targets of gender agreement. Pashai, at the western extreme, shows a diverse picture when it comes to gender pervasiveness. As mentioned before, gender may be lost altogether in some varieties at the western periphery of Pashai; whereas in e.g. SE Pashai, where direct object agreement in parts of the paradigm co-occurs with subject agreement in transitive clauses, such distinctions are frequently displayed also for inanimates. The grammatical pervasiveness of animacy-based gender is nowhere near the pervasiveness of sex-based gender, and its targets are almost invariably restricted to copula verbs and auxiliaries. The (split-)ergative pattern with object agreement in SE Pashai is possibly a factor that may point to a higher frequency of actual and potential contrasts in animacy being expressed than in the solidly accusative languages Khowar and Kalasha.

Third, when it comes to assignment criteria, the usual pattern for the sex-based systems is one of straightforward semantic assignment for humans and higher animates, and a combination of various factors (semantic, morphological and phonological) involved in the assignment of gender for lower animates and inanimates. In the animacy-based systems or sub-systems, geographically almost exclusively found at the western end of the region, semantics is the sole criterion. It also seems likely that a shift from largely non-semantic gender, such as the one in most of the Indo-Aryan languages, to largely semantic gender, is taking place in Dameli (and possibly also in Shumashti).

As already noted, speakers of Hindu Kush Indo-Aryan languages are and have been in contact with speakers of a number of other languages spoken in the region. Let us therefore take a look at these other languages and genera, in order to relate the above findings to areality beyond Indo-Aryan.

**Other Indo-Aryan languages.** In all four of the region's non-Hindu Kush Indo-Aryan languages (Hindko, Pahari-Pothwari, Gojri and Domaaki), we find a sex-based two-term system typical of Indo-Aryan (Rehman & Robinson 2011; Weinreich 2011; Kogan 2011; Losey 2002: 105–201). Apart from the obvious semantic assignment of humans and other higher animates according to biological sex, lower animates and inanimates are found in the masculine and feminine classes alike. Like in many of the HKIA languages, at a minimum, a sub-set of nouns have overt phonological markers; and at least in Gojri and Domaaki, there is a certain co-variation between gender and declensional class membership. All four languages display gender agreement with adjectives and verbs, and in addition adnominal demonstratives agree in gender in Gojri and Domaaki, and possessives in Gojri. Only Gojri shows evidence of pronominal differentiation. There are no targets of any non-sex-based agreement in any of these languages, and no observed pronominal differentiation related to animacy.

These languages are (apart from the small Domaaki enclave in the far North) mainly spoken in the southeastern part of the region, and conform in all major aspects to the pervasive sex-based gender patterns found in the HKIA languages in the same part of the region, i.e. Kashmiri and various Shina and Kohistani varieties. It is fair to assume a high level of prolonged language contact between at least Kashmiri and one or more of the languages of the Punjabi continuum, whether known as Pahari, Pothwari or Hindko, and possibly also between some of the eastern Kohistani languages and Hindko. However, in most of the areas where there is some overlap between speakers of HKIA and speakers of other Indo-Aryan languages, there is no clear dominance relationship, perhaps with Hindko-dominated parts of Pakistan-held Kashmir as an exception (Rehman 2011: 219). Both Gojri and Domaaki are examples of low-status languages vis-à-vis almost any other language communities that they have been in contact with (Losey 2002: 2–4; Weinreich 1999), and in spite of some intra-regional variations related to the relative socioeconomic status of the Gujar community (Hallberg & O'Leary 1992: 98–99, 143–144), there is no evidence of any significant influence exerted by Gojri on any of the HKIA languages.

**Iranian languages.** Iranian languages are predominantly found in the western half of the outlined region. They belong to different groupings, and their presence, and relative influence, in the area are of very different time depths. Of the nine Iranian languages represented, only three – Pashto, Shughni and Munji/Yidgha – display a sex-based gender system of some kind (Bashir 2009; Èdel'man & Dodykhudoeva 2009a,b; Kieffer 2003; 2009; Morgenstierne 1938: 110–167; Robson & Tegey 2009; Skjaervø 1989; Windfuhr & Perry 2009). In Munji/Yidgha, gender as a whole is probably in radical decline. In Shughni, the gender categories

show evidence of having restructured as to form a system of semantic classes rather than primarily being assigned on the basis of sex. Only in Pashto, which is also the language in the closest long-time contact with Indo-Aryan, do we find a two-term system akin to the typical Indo-Aryan one, with adjectives, verbs and adnominal demonstratives as agreement targets, and a certain co-variation between gender and declensional membership. Pashto and Shughni are the only Iranian languages in the sample that express pronominal gender. The rest of the Iranian languages of the region have long since lost the sex-based gender systems (masculine–feminine–neuter and masculine–feminine) that characterised their proto-languages (Skjaervø 2009b: 71; Skjaervø 2009a: 204; Yoshida 2009: 288; Durkin-Meisterernst 2009: 242–243). Although animacy distinctions are not part of agreement morphology, animacy does play a role in various forms of Persian, as certain plural allomorphs are found almost exclusively with animate nouns (Windfuhr & Perry 2009: 431), and animacy or humanness, along with register, also governs pronominal choices (2009: 435).

It is notable that it is exactly in the transitional area between Iranian and Indo-Aryan, i.e. in the western-most part of the region, that we find both a number of Iranian languages without gender, and those HKIA languages and dialects that have either lost sex-based gender altogether, or are in the process of shifting away from a primarily sex-based system to a system where animacy distinctions are becoming grammaticalized alongside an existing sex-based system. The gender-reduced systems are found primarily in the northwest, and the systems with overlapping sex-based gender and animacy in the southwest. There is possibly a correlation between gender-preserving Pashto being the most influential language of wider communication in the southwest and the retention of a masculine–feminine contrast in e.g. most Pashai and Kunar varieties. This is in contrast with the Chitral languages, which show evidence, in many parts of their language systems, of long-standing and far-reaching contact with gender-reduced Iranian languages in particular, and with a larger Central Asian contact zone in a more general sense (Bashir 1996: 176–177). Of particular interest is the now historical but crucial contact between speakers of HKIA Khowar and Iranian Wakhi. While Wakhi of today is the less influential of the two in areas where they overlap, the relationship was most likely of a symmetrical kind in a remote past, as evidenced in cross-borrowing of basic vocabulary (Morgenstierne 1936; Morgenstierne 1938: 441–442; Bashir 2007: 208–210). Different varieties of gender-less Persian, whether literary Persian, Dari or Tajik, have also had a significant (and recent) impact on the languages of Chitral and adjacent areas across the Afghanistan border in the northwestern corner of the Hindu Kush region, as a learned language and a *lingua franca*.

**Nuristani languages.** In three of the five Nuristani languages we find a two-term system of the Indo-Aryan type: in Waigali (Degener 1998: 39–91), Ashkun (Morgenstierne 1929; Morgenstierne 1934a; Morgenstierne 1952; Buddruss 2006; Grjunberg 1999) and Kati/Kamviri (Strand 2015; Èdel'man 1983: 59–71), whereas its presence in Prasun is doubtful (Morgenstierne 1949; Buddruss & Degener 2017: 69). The available data for the remaining language, Tregami, is insufficient to draw any conclusions (Morgenstierne 1952). Only Kati/Kamviri displays pronominal gender differentiation.

Although there is evidence for Nuristan and the Nuristani languages as an ancient centre of small-scale diffusion (Liljegren & Svärd 2017), Nuristani stands in most aspects, especially in more recent times, at the receiving end of contact-induced change, especially from Iranian Pashto and Persian (Degener 2002: 103). As far as gender is concerned, the possible erosion in Prasun may be attributable to the same areal influences from adjacent and influential gender-deprived Iranian languages, as was already suggested above in regard to the HKIA Chitral languages.

**Turkic languages.** There is a general absence of gender distinctions in Turkic languages, whether as overt markers of nouns or as an agreement feature (Kornfilt 2009: 530). Neither are there any pronominal distinctions in these languages. This is equally true of the two Turkic languages, Uzbek (Boeschoten 1998) and Kirghiz (Kirchner 1998), spoken by populations at the northern periphery of the Hindu Kush region.

There is no present-day overlap, or at best marginally so, between any of the HKIA communities and any of the relatively nearby Turkic-speaking groups. However, it has been suggested that at least the northern-most fringes of the Hindu Kush, together with the Pamirs and perhaps a larger region to the North, form a contact area (Èdel'man 1980; Payne 1989: 423), or alternatively a transit zone between South and Central Asia (Tikkanen 2008: 253), and it is not wholly farfetched to consider Turkic as a component of it. Bashir (1988: 402–421) points out several grammatical features (e.g. inferentiality), primarily in Kalasha and Khowar, with Turkic as their ultimate source, either mediated by certain Iranian Pamir languages or the result of a Turkic substrate. Besides, as Johanson (2013: 104) remarks, the role of Turkic in the massive gender loss in Iranian at large is yet to be fully explored.

**Tibeto-Burman languages.** Similar to what was said about Turkic, gender in its canonical sense is not a feature generally present in Tibeto-Burman. That is also largely true of Purik, a Tibeto-Burman language spoken at the south-eastern periphery of the region, although there are traces of derivational morphemes indicating male or female sex (Zemp 2013: 118–127). In closely related



Balti (Bielmeier 1985: 81; Read 1934: 4), the other Tibeto-Burman language represented in the region, we find to a larger extent such markers, postposed to some nouns denoting humans or other animates, signalling the sex of the person or animal referred to: *po* or *pho* for male, and *mo* or *ngo* for female (see §6 for formally and functionally similar markers in Brokskat). This type of sex marking or gender marking on the nouns themselves, without any reflexes in agreement patterns, should not be confused with grammatical gender as we have defined it here. In the same vein, an entirely semantically transparent pronominal differentiation can be made in Balti between human male, *kho*, human female, *mo*, and everything else (or when the sex is unknown), *do* (Read 1934: 12–13; Bielmeier 1985: 76).

It is primarily the Shina languages in the East that show traces of interaction with Tibeto-Burman (unless we, along with Tikkanen 1988: 305, consider the possibility that some of the peculiarities of Kashmiri vis-à-vis other Indo-Aryan languages might be attributed to an ancient Proto-Tibetan or Sinitic substrate). Presently, only some groups of speakers of Gilgiti Shina type varieties in Baltistan and the Brokskat community can be said to stand in any such direct and significant contact relationships, and it is only in the latter case that Tibetan plays the role of an influential donor language. It seems likely that the relationship has been more symmetrical in the past; alternatively, we would have to assume a major Tibetan substrate in the eastern Shina-speaking area. That would for instance explain agent-marking (as well as some of its formal reflexes) in Gilgiti as well as in Kohistani Shina (Liljegren 2014: 162–163; Bailey 1924: 211; Hook & Koul 2004: 213–214). In the gender domain, however, Tibeto-Burman contacts do not seem to have led to any loss or restructuring in adjacent HKIA languages, although we lack substantial information on gender assignment in Tibetan loan vocabulary in Brokskat. The continued (and perhaps strengthened) use of overt sex-marking for higher animates in Balti, and not in Purik, seems to point to Shina influences on Balti, and not the other way around.

**Burushaski.** In the northern part of the region, in close proximity to Indo-Aryan Shina, Indo-Aryan Khowar and Iranian Wakhi, the language isolate Burushaski is spoken. Burushaski has four genders, which makes it the language with the largest number of genders in the entire region. Although the number of differentiating values differs greatly from one part of the grammar to another, or from one target to another (including demonstratives, numerals, verbs, possessives and to some extent adjectives), there is a maximum four-way differentiation between human masculine (HM), human feminine (HF), and two non-human categories that traditionally have been given the labels *x* and *y* (Willson 1996: 8–9; Berger 1998: 33–34). Somewhat simplified HM is human male, HF is human female,

x is non-human animate, and y inanimate. However, in reality the relationship between the genders x and y is not quite as straightforwardly related to animacy; x includes not only animals but also fruit and some other count nouns, whereas y is the gender of abstract notions and mass nouns, but also includes e.g. trees and buildings (Yoshioka 2012: 32–33). Burushaski displays verbal agreement in gender and number with the subject as well as with the direct object of transitive clauses, as can be seen in example (17), the first by means of a suffix and the latter by means of a prefix.

- (17) Burushaski (Willson 1996: 17)  
*hilés-e dasín-mo r toofá-muts píiš ó-t-imi*  
 boy-ERG girl-OBL.F to gift(x)-PL.ABS present 3PL.X-do-3SG.HM.PST  
 ‘The boy presented gifts to the girl.’

Gender is also pronominal, but in that case HM and HF are normally neutralised, whereas x and y both have distinct forms of pronominally used demonstratives (Berger 1998: 81–82).

As Burushaski represents one of the oldest, possibly the very oldest surviving, linguistic layer in the Hindu Kush region,<sup>3</sup> it is particularly interesting from an areal point of view. While occupying a very modest territory today, the precursor of Burushaski, or other languages perhaps (but not necessarily) closely related to Burushaski, in all likelihood had a wider geographical scope before the advent of Indo-Iranian languages. It has been suggested that such substratal influence underlies some features found across Iranian, Indo-Aryan and Burushaski (Tikkanen 1988; 1999; Bashir 1988: 408–420; Êdel’man 1980). Bashir in particular attributes the gender development in the Chitral languages to Burushaski rather than to Iranian, emphasizing the emergence of animacy-based contrasts. Along the same lines, Payne (1989: 423), mainly referring to Êdel’man’s proposed convergence area, attributes the shift from formal-semantic to “purely” semantic assignment in Iranian Pamir languages to a substratum related to or similar to Burushaski, with special reference to a strikingly similar four-way differentiation in Iranian Yazghulami (female human, male human, animal and inanimate), a language situated in today’s Tajikistan, only marginally outside the Hindu Kush region as defined here.

<sup>3</sup>As pointed out to me by Johanna Nichols (p.c.), this makes perfect sense in terms of linguistic geography: a language isolated along different rivers at the highest inhabitable level is almost certainly the earlier one in and has been cut off in its former lower reaches by uphill spreads of other languages.

## 10 Conclusions

We are now in a position to summarise and draw some overall conclusions regarding the presence and distribution (geographically and subclassification-wise) of various gender properties in Hindu Kush Indo-Aryan (see Figure 3).

There are two types of gender systems in the HKIA languages. A fairly typical New Indo-Aryan sex-based two-gender system is present in the majority of the HKIA languages, and in five of the six subgroups. However, it is curiously missing altogether in the two Chitral group languages, Khowar and Kalasha, both spoken in the northwestern corner of the region. Here, instead, a two-way animacy-based gender differentiation is in place. Furthermore, these two types of gender systems are combined in another few HKIA languages, all of them found in the same part of the larger region, more or less adjacent to the Chitral languages. In one of the latter languages, Dameli, the inherited sex-based gender system is most likely subject to an ongoing process of erosion, and grammaticalized animacy-distinctions have emerged, although largely in complementary distribution with remaining sex-differentiation. In many of the varieties of Pashai, the western-most extension of HKIA, an animate–inanimate differentiation serves as a sub-gender distinction within the main masculine–feminine division.

As for the entrenchment of gender, we observed important differences between the sub-groups, forming a slight decline in pervasiveness moving from East to West. However, there is also a correlation between the presence of object agreement and the reinforcement of formal gender assignment (particularly applicable to inanimate nouns), with object-agreeing languages clustering in the South, while such HKIA languages are lacking altogether in the North. As for the pervasiveness of animacy-based gender, it was similarly suggested that its functional load is higher in systems with ergative verbal alignment (such as in Pashai) than in those with a purely accusative system (such as in the Chitral group), the latter a subject for more refined, preferably corpus-based, studies. Sex-based pronominal gender is a typical Eastern feature, exclusive to Kashmiri and the Shina group, whereas the evidence for animacy-based pronominal gender is scanty and does not allow for any further generalizations.

The weight that different assignment criteria have varies from language to language, and is a topic for which more detailed language-specific studies are needed. At a general level, there is a correlation between primarily sex-based gender and semantic-formal assignment criteria on the one hand; and a correlation between animacy-based gender and more straightforward semantic assignment criteria on the other. While gender in Indo-Aryan in general often involves

declensional differences (Masica 1991: 219), this is not a general tendency in the HKIA languages.

As far as overall complexity is concerned, a few of the HKIA languages stand out, either as being of higher than average complexity or of lower than average complexity. Languages of the first kind are primarily found in the south-westernmost part of the region; these are a handful of languages in which sex-based and animacy-based gender overlap while their targets remain largely distinct. In a single language, Kashmiri, spoken in the south-easternmost part of the region, high complexity is instead related to a high number of target domains. The languages of the second kind are those two (Kalasha and Khowar) in which gender is exclusively animacy-based, and another language (Grangali) in which agreement has been reduced to a single target domain.

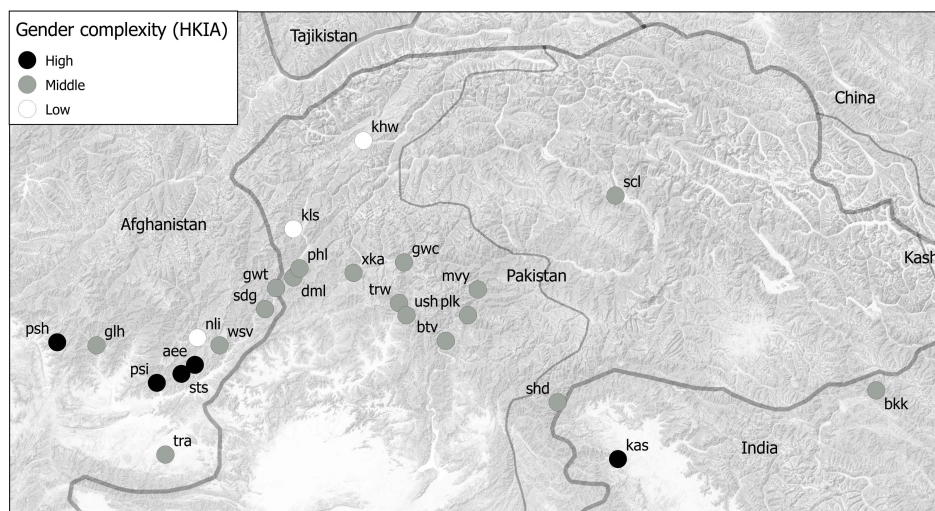


Figure 3: Gender complexity in HKIA languages

The geographical distribution of gender properties within HKIA is clearly parallel to cross-genera distribution within the region. Adjacent to the main (non-HK) Indo-Aryan continua to the Southeast as well as to Pashto, one of the more important gender-preserving Iranian languages, in the South, is where we find the most pervasive sex-based gender systems in HKIA. At the other end, i.e. the Northwest, the gender-less or gender-reduced HKIA languages border with the larger Iranian-dominated region of West and Central Asia, where sex-based gender is a rare or eroding feature, in its turn adjacent to the Turkic belt of inner Asia where gender is altogether lacking. This patterning is clearly in line with Nichols'

(2003: 303) characterization of gender as a stable feature, but only as long as related languages with inherited gender are geographically clustered. We can thus expect to find that languages that have lost this feature are indeed neighbours of one another or are surrounded by non-related languages. This makes sense if we consider Morgenstierne's (1932: 51) hypothesis that the common ancestor of the two "sex-less" languages Khowar and Kalasha represents the earliest northward migration of Indo-Aryans into this region. For a prolonged period this language must have been a relatively minor component in an area where non-Indo-Aryan (perhaps Burushaski-related, or now entirely lost) languages dominated (Tikkanen 1988; Parpola 2002: 92–94), at the time isolated from the rest of the Indo-Aryan varieties from which today's HKIA languages derive. It is also fair to assume that groups of speakers of some of those other languages shifted to a Khowar-Kalasha-type language once it became a more influential element in its new environment.

Perhaps, but not necessarily, related to this is the presence of animacy-based or other semantically highly transparent gender in the North and Northwest, with Burushaski being an obvious example. While animacy-based lexical differentiation with areal manifestation very well could be the result of borrowing, it is harder to imagine such a scenario for the copula or auxiliary agreement patterns in Shumashti and in the Chitral and Pashai languages (the forms themselves also reflecting a common source); instead we have to posit either very old substratal effects, or an internal development reinforced by similar differentiations already in place in neighbouring, and at the time influential, languages. The Dameli inanimate copula form is interesting as it bears no resemblance to the forms in the other HKIA languages (cf. examples (2), (3), (12) and (13)); instead it seems to have been recruited from inherited vocabulary (Morgenstierne 1942: 138). This topic, however, deserves a great deal of more detailed research, also taking data from the Pamir region (to the North of the Hindu Kush) into account.

## Acknowledgements

I would like to thank Alla-ud-din, Bahrain, Swat, for his help with digitizing questionnaire data; Noa Lange, Stockholm, for assistance in processing and annotating audio and video recordings; and Maria Koptjevskaja Tamm and Ekaterina Melin, Stockholm, for making the contents of publications in Russian available to me. I am also thankful for the many helpful suggestions offered by Johanna Nichols, the volume editors and four anonymous peer-reviewers.

A special thanks to a number of speakers of various HKIA languages participating in four collaborative elicitation workshops organized in Islamabad and Kabul in the time period 2015 to 2017, thus contributing valuable first-hand data.

This work is part of the project *Language contact and relatedness in the Hindu-kush region*, supported by the Swedish Research Council (421-2014-631).

## Special abbreviations

The following abbreviations are not found in the Leipzig Glossing Rules:

ACT	active	REM	remote
AN	animate	SPC	specific
CV	converb	STV	stative
INAN	inanimate	TRZ	transitivizing suffix
H	human	VIS	visible
PTC	participle	X	class x (gender in Burushaski)

## References

- Audring, Jenny. 2014. Gender as a complex feature. *Language Sciences* 43. 5–17.
- Baart, Joan L. G. 1997. *The sounds and tones of Kalam Kohistan: With wordlist and texts* (Studies in Languages of Northern Pakistan 1). Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University & Summer Institute of Linguistics.
- Baart, Joan L. G. 1999. *A sketch of Kalam Kohistani grammar* (Studies in Languages of Northern Pakistan 5). Islamabad: National Institute of Pakistan Studies & Summer Institute of Linguistics.
- Baart, Joan L. G. 2014. Tone and stress in North-West Indo-Aryan: A survey. In Johanneke Caspers, Yiya Chen, Willemijn Heeren, Jos Pacilly, Niels O. Schiller & Ellen van Zanten (eds.), *Above and beyond the segments*, 1–13. Amsterdam: John Benjamins.
- Baart, Joan L. G. & Muhammad Zaman Sagar. 2004. *Kalam Kohistani texts* (Studies in Languages of Northern Pakistan 9). Islamabad: National Institute of Pakistan Studies & Summer Institute of Linguistics.
- Bailey, T. Grahame. 1924. *Grammar of the Shina (ṣiṇā) language* (Prize Publication Fund 8). London: The Royal Asiatic Society.
- Bashir, Elena L. 1988. *Topics in Kalasha syntax: An areal and typological perspective*. University of Michigan dissertation.

- Bashir, Elena L. 1996. The areal position of Khowar: South Asian and other affinities. In Elena Bashir & Israr-ud-Din (eds.), *Proceedings of the Second International Hindukush Cultural Conference* (Hindukush and Karakoram studies 1), 167–179. Karachi: Oxford University Press.
- Bashir, Elena L. 2003. Dardic. In George Cardona & Danesh Jain (eds.), *The Indo-Aryan Languages*, 818–894. London: Routledge.
- Bashir, Elena L. 2007. Contact-induced change in Khowar. In Heather Bolton & Saeed Shafqat (eds.), *New perspectives on Pakistan: Visions for the future*, 205–238. Oxford: Oxford University Press.
- Bashir, Elena L. 2009. Wakhi. In Gernot Windfuhr (ed.), *The Iranian languages*, 825–862. London: Routledge.
- Berger, Hermann. 1998. *Die Burushaski-Sprache von Hunza und Nager 1. Grammatik* (Neuindische Studien 13). Wiesbaden: Harrassowitz.
- Bielmeier, Roland. 1985. *Das Märchen vom Prinzen ČobzaŇ: Eine tibetische Erzählung aus Baltistan: Text, Übersetzung, Grammatik und westtibetisch vergleichendes Glossar* (Beiträge zur tibetischen Erzählforschung 6). Sankt Augustin: VGH Wissenschaftsverlag.
- Boeschoten, Hendrik. 1998. Uzbek. In Lars Johanson & Éva Á. Csató (eds.), *The Turkic languages*, Reprint edition, 357–378. London: Routledge.
- Buddruss, Georg. 1960. *Die Sprache von Wotapur und Katarqala: Linguistische Studien im afghanischen Hindukusch* (Bonner Orientalische Studien 9). Bonn: Selbstverlag des Orientalischen Seminars der Universität Bonn.
- Buddruss, Georg. 1967. *Die Sprache von Sau in Ostafghanistan: Beiträge zur Kenntnis des dardischen Phalūra* (Münchener Studien zur Sprachwissenschaft Beiheft [Supplement] M). Munich: Kitzinger in Kommission.
- Buddruss, Georg. 2006. Drei Texte in der Wama-Sprache des afghanischen Hindukusch. In Mikhail N. Bogolyubov (ed.), *Indoiranskoe yazykoznanie i tipologiya yazykovykh situatsii, sbornik statei k 75-letiyu professora A. L. Gryunberga (1930–1995)*, 177–200. St. Petersburg: Nauka.
- Buddruss, Georg & Almuth Degener. 2017. *Materialien zur Prasun-Sprache des afghanischen Hindukusch 2. Grammatik* (Harvard Oriental Series 84). Cambridge: Harvard University Press.
- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2013a. Number of genders. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/30>.
- Corbett, Greville G. 2013b. Sex-based and non-sex-based gender systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of*

- language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/31>.
- Corbett, Greville G. 2013c. 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. <http://wals.info/chapter/32>.
- Corbett, Greville G. 2014. Gender typology. In Greville G. Corbett (ed.), *The expression of gender*, 87–130. Berlin: Mouton De Gruyter.
- Dahl, Östen. 2000. Elementary gender distinctions. In Matti Rissanen, Terttu Nevalainen & Mirja Saari (eds.), *Gender in grammar and cognition*. Vol. 2: *Elementary gender distinctions: Manifestations of gender* (Trends in Linguistics. Studies and Monographs 124), 577–593. Berlin: de Gruyter.
- Decker, Sandra J. 1992. Ushojo. In Calvin R. Rensch, Sandra J. Decker & Daniel G. Hallberg (eds.), *Languages of Kohistan* (Sociolinguistic Survey of Northern Pakistan 1), 65–80, 193–205. Islamabad: National Institute of Pakistani Studies & Summer Institute of Linguistics.
- Degener, Almuth. 1998. *Die Sprache von Nisheygram im afghanischen Hindukusch*. Wiesbaden: Harrassowitz Verlag.
- Degener, Almuth. 2002. The Nuristani languages. In Nicholas Sims-Williams (ed.), *Indo-Iranian languages and peoples* (Proceedings of the British Academy 116), 103–117. Oxford: Published for the British Academy by Oxford University Press.
- Degener, Almuth. 2008. *Shina-Texte aus Gilgit (Nord-Pakistan): Sprichwörter und Materialien zum Volksglauben, gesammelt von Mohammad Amin Zia*. Wiesbaden: Harrassowitz.
- Di Garbo, Francesca. 2016. Exploring grammatical complexity crosslinguistically: The case of gender. *Linguistic Discovery* 14(1). 46–85.
- Durkin-Meisterernst, Desmond. 2009. Khwarezmian. In Gernot Windfuhr (ed.), *The Iranian languages*, 336–376. London: Routledge.
- Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2019. *Ethnologue: Languages of the world, twenty-second edition*. Online version. Dallas: SIL International. <http://www.ethnologue.com>, accessed 2019-5-6.
- Èdel'man, Džoi Iosifovna. 1980. K substratnomu naslediju central'no-aziatskogo jazykovogo sojuza. *Voprosy jazykoznanija* 5. 21–32.
- Èdel'man, Džoi Iosifovna. 1983. *The Dardic and Nuristani languages* (Languages of Asia and Africa). Moscow: Nauka.
- Èdel'man, Džoi Iosifovna & Leila R. Dodykhudoeva. 2009a. Shughni. In Gernot Windfuhr (ed.), *The Iranian languages*, 787–824. London: Routledge.



- Èdel'man, Džoi Iosifovna & Leila R. Dodykhudoeva. 2009b. The Pamir languages. In Gernot Windfuhr (ed.), *The Iranian languages*, 773–786. London: Routledge.
- Emeneau, Murray B. 1965. India and historical grammar. Lecture at Annamalai University in 1959. Annamalainagar.
- Fussman, Gérard. 1972. *Atlas linguistique des parlers dardes et kafirs*. Paris: École française d'Extrême-Orient.
- Greenberg, Joseph H. 1978. How does a language acquire gender markers? In Joseph H. Greenberg, Charles Ferguson & Edith Moravcsik (eds.), *Universals of human language*. Vol. 3: *How does a language acquire gender markers?: Word structure*, 47–82. Stanford: Stanford University Press.
- Grierson, George A. 1927. *Linguistic survey of India*. Vol. 1.1: *Introduction*. Calcutta: Government of India, Central Publication Branch.
- Grierson, George A. 1929. *Torwali: An account of a Dardic language of the Swat Kohistan* (Prize publications fund 11). London: Royal Asiatic society.
- Grjunberg, Aleksandr L. 1971. K dialektologii dardiskikh jazykov (glangali i zemi-aki). In N. A. Dvoriankov (ed.), *Indijskaja i iranskaja filologija: Voprosy dialektologii*, 3–29. Moscow: Nauka.
- Grjunberg, Aleksandr L. 1999. Ashkun jazyk. In Džoi Iosifovna Èdel'man (ed.), *Jazyki mira. Dardskie i nuristanskie jazyki*, 126–131. Moscow: Indrik.
- Hallberg, Calinda E & Clare F. O'Leary. 1992. Dialect variation and multilingualism among Gujars in Pakistan. In Calvin R. Rensch, Calinda E. Hallberg & Clare F. O'Leary (eds.), *Hindko and Gujari* (Sociolinguistic Survey of Northern Pakistan 3), 91–194, 259–302. Islamabad: National Institute of Pakistan Studies & Summer Institute of Linguistics.
- Hallberg, Daniel G. & Calinda E. Hallberg. 1999. *Indus Kohistani: A preliminary phonological and morphological analysis* (Studies in Languages of Northern Pakistan 8). Islamabad: National Institute of Pakistan Studies & Summer Institute of Linguistics.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath (eds.). 2018. *Glottolog 3.2*. Jena: Max Planck Institute for the Science of Human History. <http://glottolog.org/>.
- Heegård Petersen, Jan. 2006. *Local case-marking in Kalasha*. University of Copenhagen dissertation.
- Heegård Petersen, Jan. 2015. Kalasha texts – with introductory grammar. *Acta Linguistica Hafniensia* 47(Supplement 1). 1–275.
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.
- Hook, Peter Edwin & Omkar N. Koul. 2004. Case as agreement: Non-nominative subjects in Eastern Shina, non-dative objects in Kashmiri and Poguli, and labile subjects in Kashmiri and Gujarai intransitive inceptives. In Peri Bhaskararao

- & Karumuri Venkata Subbarao (eds.), *Non-nominative subjects*, vol. 1, 213–225. Amsterdam: John Benjamins.
- Johanson, Lars. 2013. *Structural factors in Turkic language contacts*. London: Routledge.
- Kieffer, Charles M. 2003. *Hazara iv. Hazaragi dialect* – *Encyclopaedia Iranica*. <http://www.iranicaonline.org/articles/hazara-4>, accessed 2014-5-14.
- Kieffer, Charles M. 2009. Parachi. In Gernot Windfuhr (ed.), *The Iranian languages*, 693–720. London: Routledge.
- Kirchner, Mark. 1998. Kirghiz. In Lars Johanson & Éva Á. Csató (eds.), *The Turkic languages*, Reprint edition, 344–356. London: Routledge.
- Klaiman, Miriam H. 1987. Mechanisms of ergativity in South Asia. *Lingua* 71(1–4). 61–102. DOI:10.1016/0024-3841(87)90068-4
- Kogan, A. I. 2011. Pothohari jazyk. In Tatiana. I. Oranskaia, Yulia V. Mazurova, Andrej A. Kibrik, Leonid I. Kulikov & Aleksandr Y. Rusakov (eds.), *Jazyki mira: Novye indoariiskie jazyki*, 516–527. Moskva: Academia.
- Kornfilt, Jaklin. 2009. Turkish and the Turkic languages. In Bernard Comrie (ed.), *The world's major languages*, 519–544. London: Routledge.
- Koul, Omkar N. 2003. Kashmiri. In George Cardona & Danesh Jain (eds.), *The Indo-Aryan languages*, 895–952. London: Routledge.
- Lehr, Rachel. 2014. *A descriptive grammar of Pashai: The language and speech community of Darrai Nur*. Chicago: University of Chicago PhD Dissertation.
- Liljegren, Henrik. 2009. The Dangari tongue of Choke and Machoke: Tracing the proto-language of Shina enclaves in the Hindu Kush. *Acta Orientalia* 70. 7–62.
- Liljegren, Henrik. 2013. Notes on Kalkoti: A Shina language with strong Kohistani influences. *Linguistic Discovery* 11(1). 129–160. DOI:10.1349/PS1.1537-0852.A.423
- Liljegren, Henrik. 2014. A survey of alignment features in the Greater Hindukush with special references to Indo-Aryan. In Pirkko Suihkonen & Lindsay J. Whaley (eds.), *On diversity and complexity of languages spoken in Europe and North and Central Asia* (Studies in Language Companion Series 164), 133–174. Amsterdam: John Benjamins.
- Liljegren, Henrik. 2016. *A grammar of Palula* (Studies in Diversity Linguistics 8). Berlin: Language Science Press.
- Liljegren, Henrik. 2017. Profiling Indo-Aryan in the Hindukush-Karakoram: A preliminary study of micro-typological patterns. *Journal of South Asian Languages and Linguistics* 4(1). 107–156. DOI:10.1515/jsall-2017-0004
- Liljegren, Henrik & Erik Svärd. 2017. Bisyndetic contrast marking in the Hindukush: Additional evidence of a historical contact zone. *Journal of Language Contact* 10(3). 450–484.

- Losey, Wayne E. 2002. *Writing Gojri: Linguistic and sociolinguistic constraints on a standardized orthography for the Gujars of South Asia*. University of North Dakota MA thesis.
- Lubberger, Beate. 2014. *A description and analysis of four metarepresentation markers of Indus Kohistani*. Grand Forks, North Dakota: University of North Dakota MA thesis.
- Lunsford, Wayne A. 2001. *An overview of linguistic structures in Torwali: A language of northern Pakistan*. University of Texas at Arlington MA thesis.
- Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Masica, Colin P. 2001. The definition and significance of linguistic areas: Methods, pitfalls, and possibilities (with special reference to the validity of South Asia as a linguistic area). In Peri Bhaskararao (ed.), *The yearbook of South Asian languages and linguistics 2001*, 205–267. London: SAGE.
- Morgenstierne, Georg. 1929. The language of the Ashkun Kafirs. *Norsk Tidsskrift for Sprogvidenskap* 2. 192–289.
- Morgenstierne, Georg. 1932. *Report on a linguistic mission to North-Western India* (Serie C III - 1). Oslo: Instituttet for sammenlignende kulturforskning.
- Morgenstierne, Georg. 1934a. Additional notes on Ashkun. *Norsk Tidsskrift for Sprogvidenskap* 7. 56–115.
- Morgenstierne, Georg. 1934b. Notes on Tirahi. *Acta Orientalia* 12. 161–189.
- Morgenstierne, Georg. 1936. Iranian elements in Khowar. *Bulletin of the School of Oriental Studies, University of London* 8(2/3). 657–671.
- Morgenstierne, Georg. 1938. *Indo-Iranian frontier languages*. Vol. 2: *Iranian Pamir languages (Yidgha-Munji, Sanglechi-Ishkashmi and Wakhi)*. Oslo: Universitetsforlaget.
- Morgenstierne, Georg. 1942. Notes on Dameli: A Kafir-Dardic language of Chitral. *Norsk Tidsskrift for Sprogvidenskap* 12. 115–198.
- Morgenstierne, Georg. 1945. Notes on Shumashti: A Dardic dialect of the Gawar-Bati type. *Norsk Tidsskrift for Sprogvidenskap* 13. 239–281.
- Morgenstierne, Georg. 1949. The language of the Prasun Kafirs. *Norsk Tidsskrift for Sprogvidenskap* 15. 186–334.
- Morgenstierne, Georg. 1950. *Notes on Gawar-Bati*. Oslo: Det Norske Videnskaps-Akademi.
- Morgenstierne, Georg. 1952. Linguistic gleanings from Nuristan. *Norsk Tidsskrift for Sprogvidenskap* 16. 117–135.
- Morgenstierne, Georg. 1967. *Indo-Iranian frontier languages*. Vol. 3: *The Pashai language, 1, Grammar*. Oslo: Universitetsforlaget.

- Morgenstierne, Georg. 1974. Languages of Nuristan and surrounding regions. In *Cultures of the Hindukush: Selected papers from the Hindu-Kush Cultural Conference held at Moesgård 1970*, vol. 1 (Beiträge zur Südasienforschung), 1–10. Wiesbaden: Steiner.
- Nichols, Johanna. 2003. Diversity and stability in language. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 283–310. Oxford: Blackwell.
- Parpola, Asko. 2002. From the dialects of Old Indo-Aryan to Proto-Indo-Aryan and Proto-Iranian. In Nicholas Sims-Williams (ed.), *Indo-Iranian languages and peoples* (Proceedings of the British Academy 116), 43–102. Oxford: Published for the British Academy by Oxford University Press.
- Payne, John R. 1989. Pamir languages. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 417–444. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Perder, Emil. 2013. *A grammatical description of Dameli*. Stockholm University PhD Dissertation.
- Pirejko, L. A. 1979. On the genesis of the ergative construction in Indo-Iranian. In Frans Plank (ed.), *Ergativity: Towards a theory of grammatical relations*, 481–488. London: Academic Press.
- Radloff, Carla F. & Shakil Ahmad Shakil. 1998. *Folktales in the Shina of Gilgit: Text, grammatical analysis and commentary* (Studies in Languages of Northern Pakistan 2). Islamabad: National Institute of Pakistan Studies & Summer Institute of Linguistics.
- Ramaswami, N. 1982. *Brokskat grammar* (CIIL grammar series 8). Mysore: CIIL.
- Read, Alfred F. C. 1934. *Balti grammar*. London: James G. Forlong Fund.
- Rehman, Khawaja A. 2011. Ergativity in Kundal Shahi, Kashmiri and Hindko. In Mark Turin & Bettina Zeisler (eds.), *Himalayan languages and linguistics: Studies in phonology, semantics, morphology and syntax*, 219–233. Leiden: Brill Academic Publishers.
- Rehman, Khawaja A. & Joan L. G. Baart. 2005. A first look at the language of Kundal Shahi in Azad Kashmir. *SIL Electronic Working Papers* 2005-008.
- Rehman, Khawaja A. & Mark A. Robinson. 2011. Khindko jazyk. In Tatiana. I. Oranskaia, Yulia V. Mazurova, Andrej A. Kibrik, Leonid I. Kulikov & Aleksandr Y. Rusakov (eds.), *Jazyki mira: Novye indoariiskie jazyki*, 527–537. Moskva: Academia.
- Robson, Barbara & Habibullah Tegey. 2009. Pashto. In Gernot Windfuhr (ed.), *The Iranian languages*, 721–772. London: Routledge.
- Schmidt, Ruth Laila & Razwal Kohistani. 2008. *A grammar of the Shina language of Indus Kohistan* (Beiträge zur Kenntnis südasiatischer Sprachen und Literaturen 17). Wiesbaden: Harrassowitz.

- Sharma, Devidatta. 1998. *Tribal languages of Ladakh. Part One: A concise grammar and dictionary of Brok-skad* (Studies in Tibeto-Himalayan languages 6). New Delhi: Mittal Publications.
- Siewierska, Anna. 2013. Gender distinctions in independent personal pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/44>.
- Skalmowski, Wojciech. 1985. The linguistic importance of the Dardic languages. *Journal of Central Asia* 8. 5–15.
- Skjaervø, Prods Oktor. 1989. Pashto. In Rüdiger Schmitt (ed.), *Compendium linguarum Iranicarum*, 384–410. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Skjaervø, Prods Oktor. 2009a. Middle West Iranian. In Gernot Windfuhr (ed.), *The Iranian languages*, 196–278. London: Routledge.
- Skjaervø, Prods Oktor. 2009b. Old Iranian. In Gernot Windfuhr (ed.), *The Iranian languages*, 43–195. London: Routledge.
- Strand, Richard F. 1973. Notes on the Nuristani and Dardic languages. *Journal of the American Oriental Society* 93(3). 297–305.
- Strand, Richard F. 2015. *Richard Strand's Nuristân site: Lexicon and grammar*. <http://nuristan.info/lngFrameG.html>, accessed 2016-4-13.
- Tikkanen, Bertil. 1988. On Burushaski and other ancient substrata in northwestern South Asia. *Studia Orientalia* 64. 303–325.
- Tikkanen, Bertil. 1999. Archaeological-linguistic correlations in the formation of retroflex typologies and correlating areal features in South Asia. In Roger Blench & Matthew Spriggs (eds.), *Archaeology and language IV: Language change and cultural transformation*, 138–148. London: Routledge.
- Tikkanen, Bertil. 2008. Some areal phonological isoglosses in the transit zone between South and Central Asia. In Israr-ud-Din (ed.), *Proceedings of the third International Hindu Kush Cultural Conference*, 250–262. Karachi: Oxford University Press.
- Toporov, Vladimir Nikolayevich. 1970. About the phonological typology of Burushaski. In Roman Jakobson & Shigeo Kawamoto (eds.), *Studies in general and oriental linguistics presented to Shiro Hattori on the occasion of his sixtieth birthday*, 632–647. Tokyo: TEC Corporation for Language & Educational Research.
- Verbeke, Saartje. 2013. *Alignment and ergativity in new Indo-Aryan languages*. Berlin: De Gruyter Mouton.
- Weinreich, Matthias. 1999. Der Domaaki-Dialekt von Nager. In Georg Buddruss (ed.), *Studien zur Indologie und Iranistik. Band 22*. 203–214. Reinbek: Verlag für Orientalistische Fachpublikationen.

- Weinreich, Matthias. 2011. Domaaki jazyk. In Tatiana. I. Oranskaia, Yulia V. Mazurova, Andrej A. Kibrik, Leonid I. Kulikov & Aleksandr Y. Rusakov (eds.), *Jazyki mira: Novye indoariiskie jazyki*, 165–194. Moscow: Academia.
- Willson, Stephen R. 1996. Verb agreement and case marking in Burushaski. *Work Papers of the Summer Institute of Linguistics, University of North Dakota* 40. 1–71.
- Windfuhr, Gernot & John R. Perry. 2009. Persian and Tajik. In Gernot Windfuhr (ed.), *The Iranian languages*, 416–544. London: Routledge.
- Yoshida, Yutaka. 2009. Sogdian. In Gernot Windfuhr (ed.), *The Iranian languages*, 279–335. London: Routledge.
- Yoshioka, Noboru. 2012. *A reference grammar of Eastern Burushaski*. Tokyo: Tokyo University of Foreign Studies PhD Dissertation.
- Zemp, Marius. 2013. *A historical grammar of the Tibetan dialect spoken in Kargil (Purik)*. Berne: University of Berne PhD Dissertation.

# Name index

- Abiodun, Michael A., 114  
Acquaviva, Paolo, 188  
Aikhenvald, Alexandra Y., 33, 114,  
200, 229, 244, 245, 261–263,  
268  
Allan, Edward Jay, 131, 132, 134, 135  
Alva, Elda Alicia, 41  
Arias-Trejo, Natalia, 41  
Årsjö, Britten, 229, 239, 272  
Audring, Jenny, 1, 18, 22, 25, 30, 31,  
35, 64, 307  
  
Baart, Joan L. G., 280, 286, 294, 298,  
306  
Babou, Cheikh Anta, 114  
Bailey, T. Grahame, 286, 315  
Bakker, Dik, 228  
Barlow, Michael, 34  
Bashir, Elena L., 280, 285, 286, 297,  
305, 312–314, 316  
Beam, Mary S., 155  
Bentley, Mayrene, 23  
Berger, Hermann, 315, 316  
Bewer, Franziska, 43  
Bickel, Balthasar, 67, 73, 74, 76, 79  
Bielmeier, Roland, 315  
Bittner, Dagmar, 43  
Blench, Roger, 130  
Blom, Elma, 39, 40  
Bodomo, Adams, 118–121  
Boeschoten, Hendrik, 314  
Bokula, François-Xavier, 114  
  
Bond, Oliver, 15  
Booi, Geert, 29  
Borchardt, Nadine, 107  
Brown, Dunstan, 15, 18  
Brown, Lea, 195, 229  
Bruce, Leslie P., 262  
Brugmann, Karl, 197  
Buddruss, Georg, 286, 314  
Bybee, Joan, 270  
  
Canu, Gaston, 114  
Carter, Hazel, 262  
Caselli, Maria Cristina, 41, 43, 44  
Christaller, Johann Gottlieb, 116  
Connell, Bruce A., 114  
Conrad, Robert J., 229, 239, 245, 252  
Corbett, Greville G., 1, 15, 16, 18–  
21, 23, 25, 28, 30, 31, 33, 34,  
36, 55–57, 59, 64, 67, 68, 72,  
83, 96–99, 102, 105, 106, 115,  
131–133, 139, 140, 148, 188,  
192, 200, 209, 218–221, 226,  
227, 231, 237, 254, 256, 257,  
262, 281, 282, 285, 287, 297,  
300  
Cornips, Leonie, 42, 44  
Corris, Miriam, 229  
Creissels, Denis, 83  
Cridland, Elizabeth A., 155  
Croft, William, 227  
Cysouw, Michael, 149

*Name Name index*

- Dahl, Östen, 1, 53, 55, 58–60, 64, 70,  
149, 162, 203, 289  
Dakubu, Mary Esther Kropp, 116, 121  
Dale, Rick, 54, 55  
de Wolf, Paul P., 112, 114  
Decker, Sandra J., 286  
Degener, Almuth, 286, 314  
Demuth, Katherine, 39, 41, 46  
Di Garbo, Francesca, 1, 22, 23, 28, 148,  
149, 163, 225, 227, 230, 231,  
245, 250, 253, 256, 261, 262,  
272, 307  
Dimmendaal, Gerrit, 26  
Dixon, Robert M. W., 33, 42, 150, 162,  
175  
Dobrushina, Nina, 73  
Dodykhudoeva, Leila R., 312  
Dol, Philomena Hedwig, 229, 230,  
237  
Dolphyne, Florence Abena, 116, 121  
Doneux, Jean L., 114  
Donohue, Mark, 64, 229, 264–266  
Drabbe, Peter, 200  
Dryer, Matthew S., 54, 172, 192–195,  
229, 272  
Durie, Mark, 209  
Durkin-Meisterernst, Desmond,  
313  
Eberhard, David M., 282  
Èdel'man, Džoi Iosifovna, 280, 312,  
314, 316  
Edwards-Fumey, Deborah, 207  
Eichler, Nadine, 39, 41, 43, 46  
Elugbe, Ben O., 114  
Emeneau, Murray B., 280, 281  
Enger, Hans-Olav, 18  
Essegbey, James, 137  
Evans, Nicholas, 83, 99, 200  
Fedden, Sebastian, 15, 18–21, 25, 31,  
33, 34, 72, 83, 140, 217, 218,  
228, 229, 261, 262, 267, 268  
Fiedler, Ines, 130  
Foley, William A., 78, 200, 228, 229,  
239, 245, 247, 256, 259, 263,  
268  
Frowein, Friedel Martin, 236, 250,  
253  
Fussman, Gérard, 280  
Gagliardi, Annie, 38, 42  
Garbo, Francesca Di, 22, 23, 58, 64,  
86  
Geraghty, Paul, 165  
Geurtjens, Hendrik, 205  
Good, Jeff, 114  
Greenberg, Joseph H., 147, 287  
Grierson, George A., 286  
Grjunberg, Aleksandr L., 286, 314  
Güldemann, Tom, 100–102, 114, 115,  
130, 139, 165  
Guthrie, Malcolm, 112  
Hallberg, Calinda E, 286, 312  
Hallberg, Calinda E., 286  
Hallberg, Daniel G., 286  
Hammarström, Harald, 130, 282  
Hansford, Keir Lewis, 122–126  
Haspelmath, Martin, 54, 109  
Heath, Jeffrey, 82  
Heegård Petersen, Jan, 280, 286, 297,  
305  
Heine, Bernd, 96, 106, 114, 131, 133,  
136, 137, 139, 261  
Hengeveld, Kees, 83  
Himmelmann, Nikolaus, 166  
Hockett, Charles F., 16, 148, 165, 226,  
287



- Hoel, Hanna Marie, 229, 233  
Hook, Peter Edwin, 315  
Hulk, Aafke, 42, 44  
Hyman, Larry M., 114  
  
Iemmolo, Giorgio, 227  
  
James, Wendy, 156  
Janhunnen, Juha, 66  
Johanson, Lars, 314  
Johnson, Elizabeth K., 39  
  
Karmiloff-Smith, Annette, 42  
Keij, Brigitta, 40, 44  
Kempe, Vera, 42  
Key, Harold H., 76, 77  
Kibrik, Aleksandr E., 82  
Kibrik, Alexander E., 72, 82  
Kibrik, Andrej A., 83, 85  
Kieffer, Charles M., 312  
Kilarski, Marcin, 114  
Killian, Don, 36, 59, 150, 152  
Kirchner, Mark, 314  
Klaiman, Miriam H., 292  
Kogan, A. I., 312  
Kohistani, Razwal, 286, 305  
Köpcke, Klaus-Michael, 17  
Koptjevskaja-Tamm, Maria, 216, 272  
Kornfilt, Jaklin, 314  
Koul, Omkar N., 286, 292, 299, 315  
Kulemeka, Andrew T. C., 23  
Kulick, Don, 229  
Kulikov, Leonid, 58, 87  
Kusters, Wouter, 22, 23, 38  
  
Lacroix, René, 78, 79  
Lakoff, George, 42  
Laycock, Donald C., 24  
Leer, Jeff, 79  
Lehr, Rachel, 286, 296  
  
Leufkens, Sterre, 31  
Levins, Saul, 58  
Levy, Yonata, 42  
Lidz, Jeffrey, 38, 42  
Liljegren, Henrik, 86, 280, 286, 292,  
314, 315  
Lindström, Eva, 229, 253  
Lock, Arjen, 229, 234, 260, 267, 268  
Loporcaro, Michele, 114  
Losey, Wayne E., 312  
Loughnane, Robyn, 229  
Lubberger, Beate, 286, 293  
Lunsford, Wayne A., 286  
Lupyan, Gary, 54, 55  
Luraghi, Silvia, 255, 257  
  
MacWhinney, Brian, 39, 42, 43  
Maho, Jouni, 86  
Manessy, Gabriel, 114, 127–129  
Maniscalco, Samuele, 102  
Marchese, Lynell, 114  
Marfo, Charles, 118–121  
Mariscal, Sonia, 40, 41, 43  
Masica, Colin P., 280, 281, 292, 298,  
318  
Maslova, Elena, 79–81  
Matasović, Ranko, 58, 65, 87  
Mattissen, Johanna, 75, 76  
Mchombo, Sam A., 23  
McKenzie, Parker, 78  
McWhorter, John, 1  
Meeussen, Achille E., 108  
Miehe, Gudrun, 112, 114  
Miestamo, Matti, 22, 23, 58, 64, 86,  
149  
Mills, Anne E., 39–43  
Morgenstierne, Georg, 280, 283, 286,  
289, 304, 312–314, 319

*Name Name index*

- Mosel, Ulrike, 228, 229, 236, 249, 252, 253  
Mulford, Randa, 39, 42, 43  
Nagaraja, Keralapura Shreenivasa-  
iah, 59  
Nagayama, Yukari, 82  
Neuhaus, Simon, 102  
Nichols, Johanna, 1, 25, 56, 64, 67, 71–  
74, 76, 79, 84, 96  
O’Leary, Clare F., 286, 312  
Olsson, Bruno, 198, 272  
Onishi, Masayuki, 229, 237, 243, 261,  
268  
Osam, Emmanuel Kweku, 117, 118,  
121  
Öztürk, Balkız, 78–80  
Painter, Colin, 122, 127  
Parpola, Asko, 319  
Pasch, Helma, 114  
Passer, Matthias Benjamin, 2  
Payne, John R., 314, 316  
Perder, Emil, 286, 296  
Pérez Pereira, Miguel, 42, 43  
Perkins, Revere D., 55  
Perry, John R., 312, 313  
Phillips, Colin, 239  
Pirejko, L. A., 292  
Pizzuto, Elena, 41, 43, 44  
Plaster, Keith, 42  
Pöchtrager, Markus A., 78–80  
Polinsky, Maria, 42  
Radloff, Carla F., 286  
Ramaswami, N., 286, 299  
Read, Alfred F. C., 315  
Rehman, Khawaja A., 286, 312  
Reineke, Bigitte, 127  
Robins, Robert H., 75, 76  
Robinson, Mark A., 312  
Robinson, Stuart Payton, 229, 239,  
240, 253, 265  
Robson, Barbara, 312  
Rodina, Yulia, 40, 42  
Rumsey, Alan, 263  
Ryding, Karin C., 24  
Sagar, Muhammad Zaman, 306  
Schadeberg, Thilo, 114  
Schmidt, Ruth Laila, 286, 305  
Schulze, Wolfgang, 88  
Scorza, David, 229, 242  
Shakil, Shakil Ahmad, 286  
Sharma, Devidatta, 286, 299  
Shosted, Ryan K., 64  
Siewierska, Anna, 24, 34, 282, 285  
Sinnemäki, Kaius, 22, 25, 54  
Skalmowski, Wojciech, 280  
Skjaervø, Prods Oktor, 312, 313  
Smith-Stark, Thomas Cedric, 221  
Snider, Keith, 114, 127–129  
Spriggs, Ruth, 228, 229, 236, 249, 252,  
253  
Steele, Susan, 226  
Steeman, Sander, 27  
Stern, Jan P., 112  
Stewart, John M., 116  
Strand, Richard F., 280, 314  
Stroud, Christopher, 229  
Suter, Edgar, 197, 198, 201, 202, 205–  
207, 217  
Svantesson, Jan-Olof, 66  
Svärd, Erik, 59, 228, 229, 234, 235,  
314  
Szagun, Gisela, 40, 42  
Tegey, Habibullah, 312

Tharp, Doug, 228  
Thurston, William R., 76, 77  
Tikkanen, Bertil, 280, 314–316, 319  
Toporov, Vladimir Nikolayevich,  
280  
Trudgill, Peter, 1, 53, 64, 72  
  
Usher, Timothy, 197, 198, 201, 202,  
205–207, 217  
  
Van de Velde, Mark L. O, 110  
van Heugten, Marieke, 39  
Verbeke, Saartje, 286  
Voll, Rebecca M., 114  
Voorhoeve, Jan, 112  
  
Wälchli, Bernhard, 18, 165, 216, 229,  
234, 235, 245, 257, 272  
Watkins, Laurel J., 78  
Wegener, Claudia, 27  
Wegener, Heide, 43  
Weinreich, Matthias, 283, 312  
Welmers, William E., 109  
Westergaard, Marit, 42  
Westermann, Diedrich, 96, 127  
Willson, Stephen R, 315, 316  
Wilson, William H., 77  
Windfuhr, Gernot, 312, 313  
Wogiga, Kepas, 229, 239, 245, 252  
Wood, Esther J., 79  
  
Yoshida, Yutaka, 313  
Yoshioka, Noboru, 316  
Yu, Alan C-L., 79  
  
Zaliznjak, Andrej A., 98, 220  
Zaske, Stephanie G., 137  
Zemp, Marius, 314



# Language index

- Abau, 229, 232, 234, 238, 242, 246–248, 260, 262, 265, 267–269
- Abkhaz, 78
- Adele, 130
- Adyghe, 78
- Afro-Asiatic, 24, 57, 256
- Akan, 5, 114–119, 121, 138, 140
- Akanic, 116, 121
- Akuapem, 116
- Alamblak, 262
- Algic, 75
- Algonquian, 79
- Alutor, 82
- Ama, 229, 232, 238, 239, 242, 246, 247, 269, 270, 272
- Amharic, 261
- Anêm, 77
- Anii, 130, 137
- Anim, 7, 198, 201, 205–207, 217–219, 221
- Animere, 130, 136
- Arabic, 24, 25, 156
- Arapesh, 229, 239, 257
- Arawak, 268
- Arawan, 162
- Armenian, 58, 87
- Asante, 116
- Ashkun, 285, 314
- Atlantic, 114, 115, 255
- Au, 229, 232, 238, 242, 246–248, 264, 269
- Austroasiatic, 59
- Austronesian, 7, 77, 165, 228, 229, 231, 235, 249, 250, 253–255, 270
- Avar, 64, 72, 73, 83, 84
- Avar-Andic-Tsezic, 72
- Avatime, 130
- Awutu, 122
- Baglo, 131, 133
- Balti, 285, 315
- Baltic, 87
- Bantoid, 23, 114
- Bantu, 5, 39, 41, 46, 59, 63, 73, 97, 103, 108–116, 118, 119, 136–139, 255, 256, 259, 262, 271
- Barupu, 229
- Bateri, 284–286, 288, 291, 295, 303, 309
- Benue-Congo, 112, 114
- Benue-Kwa, 105, 115, 121
- Bininj Gun-Wok, 256
- Boro, 130
- Bosnian, 220
- Brokskat, 284, 286, 288, 291, 292, 299, 300, 306, 309, 315
- Bron, 117, 118
- Bukiyip, 229, 232, 237–239, 245–247, 250, 256, 257, 264, 269, 270
- Bulgarian, 87
- Burmeso, 228, 229, 231, 232, 237–239, 242, 246, 247, 264, 265, 268, 269, 271

*Language index*

- Burushaski, 8, 280, 283, 285, 315, 316,  
319, 320
- C'lela, 115
- Cayuvava, 76, 77
- Chadic, 219
- Chenapian, 268
- Cherepon, 122
- Chichewa, 23
- Chilisso, 284, 285
- Chinese, 75
- Chitral, 8, 283–285, 288, 297, 304,  
313, 314, 316, 317, 319
- Chukchi-Kamchatkan, 82
- Chumburung, 121–124, 127–129
- Coastal Marind, 7, 197, 198, 200–202,  
204–208, 210, 212, 214–222
- Croatian, 220
- Cushitic, 102, 140, 219
- Daghestanian, 71, 72
- Dameli, 282, 284, 286, 288, 289,  
291, 295–297, 303, 304, 306,  
309–311, 317, 319
- Dardic, 280, 282
- Dari, 283, 313
- Domaaki, 283, 312
- Dompo, 122
- Dravidian, 268
- Dutch, 26, 27, 30, 31, 40, 42, 44, 46, 47
- Dwang, 122
- Dyirbal, 42
- Edoid, 114
- Enga, 263
- Engan, 263
- English, 9, 18, 41, 42, 46, 67, 83, 85, 87,  
188, 205, 208, 216, 235
- Eton, 110
- Fante, 116–118
- Fijian, 165, 166
- Finnish, 85
- Foodo, 122, 129
- French, 29, 41–43, 64, 85
- Fula, 115
- Gade, 112, 113, 115
- Gawarbatl, 284, 286, 288, 291, 295,  
303, 304, 309
- Gawri, 284, 286, 288, 291, 294, 295,  
298, 303, 306, 309
- German, 17, 18, 41–43, 83, 216, 300
- Germanic, 58, 86, 87, 205
- Ghana-Togo-Mountain, 5, 95, 114–  
116, 121, 130, 131, 136–138
- Gichode, 121, 122, 129
- Ginyanga, 122, 129
- Gojri, 283, 312
- Gonja, 121, 122, 127, 129
- Gowro, 284, 285
- Grangali, 284, 286, 288, 291, 308, 309,  
318
- Greek, 86
- Gua, 122
- Guang, 5, 114, 115, 121, 122, 126–131,  
136, 138
- Guaraní, 57
- Gule, 147
- Gunwinygic, 256
- Gur, 114, 115
- Gwama, 147
- Hazaragi, 283
- Hebrew, 24
- Heibanic, 114
- Hindko, 283, 312
- Hindu Kush Indo-Aryan, 8, 280, 282,  
285, 292, 305, 307, 310–313,  
315, 317–320

- Humono, 113
- Icelandic, 42, 43
- Igede, 114
- Igo, 130, 136
- Ikaan, 105, 106, 112, 113
- Ikposo, 130, 136
- Indic, 58, 268
- Indo-Aryan, 8, 280–283, 287, 292, 310–319
- Indo-European, 38, 57, 58, 63, 67, 70, 87, 88, 197, 256
- Indo-Iranian, 280, 316
- Ingush, 70–73, 83–85, 88
- Iranian, 8, 87, 268, 280, 283, 312–316, 318
- Ishkashimi, 283
- Italian, 16, 17, 21, 22, 24, 25, 41, 43, 44, 46
- Iwam, 268
- Jarawara, 162
- Ju, 100
- Jul’hoan, 101, 102, 255
- Ka-Togo, 130
- Kadu, 102, 140
- Kalami, 284, 286, 288, 291
- Kalasha, 8, 283, 284, 286, 288, 291, 297, 307–311, 314, 317–319
- Kalkoti, 284–286, 288, 291, 306, 309
- Kamviri, 285, 314
- Kartvelian, 78
- Kashmiri, 8, 283–286, 288, 290–292, 298, 299, 305, 306, 308–312, 315, 317, 318
- Kati, 285, 314
- Khasi, 59
- Khoe, 100
- Khoisan, 27, 57, 100
- Khoisan, Central, 261
- Khovar, 8, 283, 284, 286, 288, 289, 291, 297, 307–311, 313–315, 317–319
- Kiowa, 78
- Kiowa-Tanoan, 78
- Kirghiz, 285, 314
- Kisi, 115
- Kohistani, 284, 286, 288, 291, 293, 294, 299, 303–306, 309, 311, 312
- Kohistani, Indus, 284–286, 288, 291, 293, 295, 303, 309
- Kohistani, Kalam, 294
- Koman, 6, 36, 147
- Komo, 147
- Kplang, 122
- Krache, 122
- Krachi, 121
- Krongo, 102
- Kru, 114
- Kunar, 284, 289, 311, 313
- Kundal Shahi, 284, 286, 288, 291, 306, 309
- Kuot, 228, 229, 232, 238, 245–247, 253, 264, 269
- Kwa, 115, 130
- Kx’a, 100, 140, 255
- Laala, 115
- Larteh, 122
- Latin, 86, 269
- Laz, 78, 80
- Left May, 229, 239, 272
- Lelemi, 130–134
- Lezgian, 71
- Lithuanian, 216
- Logba, 116, 130

*Language index*

- Lower Cross, 114  
Lower Sepik, 229, 239, 257, 270  
Lower Sepik-Ramu, 78, 228, 229, 239, 256, 257  
  
Macedonian, 87  
Manambu, 229, 232, 238, 240, 242, 243, 245–247, 260, 262, 264, 269  
Mandarin, 75, 76  
Mankiyali, 284, 285  
Maybrat, 229, 230, 232, 237–239, 242, 246, 247, 264, 269  
Mbaic, 114  
Mbane, 115  
Mek, 228, 229, 234, 235, 248, 249, 257, 272  
Mende, 229, 232, 233, 238–240, 242, 243, 246, 247, 264, 269, 270  
Menya, 270  
Mian, 83, 198, 217–220, 228, 229, 232, 237, 238, 242, 246, 247, 255, 261, 264, 267, 269  
Miyobe, 115  
Mongolian, 66–70, 72  
Moore, 114  
Motuna, 229, 232, 237, 238, 242, 243, 246, 247, 261, 264, 268, 269  
Mundabli, 114  
Munji, 283, 312  
  
Na-Togo, 130, 131  
Nakh, 70, 72  
Nakh-Daghestanian, 57, 58, 64, 70, 71, 73, 83, 87, 88  
Nalca, 165, 228, 229, 232, 234, 235, 237, 238, 245–249, 257, 264, 269, 272  
Nawuri, 122  
  
Nchumbulu, 122  
Nchumburu, 121  
Ndu, 229, 243, 260, 262  
Nehan-North Bougainville, 229  
Ngala, 24  
Niger-Congo, 5, 23, 39, 56, 57, 95, 96, 101–103, 105, 109, 112–119, 121, 124, 127, 129, 133, 134, 136–140, 254, 256  
Nilotic, 26, 150  
Nivkh, 75, 76  
Nkami, 122  
Nkonya, 122, 127  
North Bougainville, 229, 239, 240, 253, 265  
Nterato, 122  
Nupoid, 112  
Nuristani, 8, 280, 285, 314  
Nyangbo, 130, 137  
  
Oceanic, 228, 229, 249, 250, 253, 254, 270  
Ok, 7, 198, 217–219, 222, 228, 229, 255, 261  
Oksapmin, 229, 232, 238, 242, 246, 247, 255, 264, 269  
Opo, 147  
Opuo, 147  
Oromo, 261  
  
Pahari-Pothwari, 283, 312  
Palai, 229  
Palula, 284, 286, 288, 291–293, 298–300, 303, 306, 309  
Parachi, 283  
Pashai, 8, 283, 284, 287, 289–291, 296, 297, 304, 308–311, 313, 317, 319  
Pashai, Northeast, 284, 286, 288, 291



- Pashai, Northwest, 284, 286, 288, 291  
Pashai, Southeast, 284, 286, 288, 296  
Pashai, Southwest, 284, 286, 288  
Pashto, 283, 285, 312–314, 318  
Persian, 87, 283, 313, 314  
Polynesian, 77  
Potou-Akanic, 95, 116, 121  
Prasun, 285, 314  
Proto-Slavic, 70, 86  
Punjabi, 283, 312  
Purik, 285, 314, 315
- Romance, 58, 87  
Romanian, 33, 101, 102  
Rotokas, 229, 232, 238–240, 242, 243,  
246, 247, 253, 261, 264, 265,  
269  
Russian, 40, 42, 66–70, 72, 83, 85, 208,  
219, 220, 256, 319
- Sandawe, 27  
Sanglechi, 283  
Savosavo, 27, 28  
Sawi, 284, 286, 288, 291, 299, 306, 309  
Scandinavian, 59  
Sekpele, 130  
Selee, 130, 137  
Semitic, 58  
Sepik, 24, 228, 229, 233, 234, 239, 240,  
248, 260, 262, 263, 268, 271  
Serbian, 220  
Sesotho, 46  
Shina, 283, 284, 286, 288, 291–294,  
299, 303–306, 309–312, 315,  
317  
Shina, Gilgiti, 284, 286, 288, 291–293,  
299, 306, 309, 315  
Shina, Kohistani, 292, 305, 315  
Shughni, 283, 312, 313
- Shumashti, 8, 284, 286, 288, 289, 291,  
296, 297, 303, 304, 308, 309,  
311, 319  
Sinitic, 315  
Siswati, 46  
Siwai, 229  
Siwu, 130  
Sko, 228, 229, 239, 260, 262, 271  
Skou, 229, 232, 238, 239, 242, 246,  
247, 260, 262, 264, 269, 271  
Slavic, 87  
Somali, 102  
South Bougainville, 229, 243, 261  
Spanish, 26, 41–43, 46, 64  
Spanish, Cantabrian, 261  
Sulka, 228  
Swahili, 59, 97–99, 103–105, 109, 110
- Tafi, 130  
Tagalog, 79, 80, 165, 166  
Taiap, 229, 232, 238, 239, 242, 246,  
247, 260, 262, 264, 269, 271  
Tajik, 283, 313  
Talodic, 114  
Tchumbuli, 122  
Teop, 228, 229, 231, 232, 235, 236, 238,  
242, 246, 247, 249, 250, 253,  
254, 264, 269, 270  
Tetemang, 132, 133  
Tibetan, 315  
Tibeto-Burman, 8, 280, 285, 314, 315  
Tirahi, 282, 284, 286, 288, 291, 309  
Tonga, 261, 262  
Torricelli, 6, 171, 186, 228, 229, 239,  
256, 257, 270  
Torwali, 284, 286, 288, 291, 295, 303,  
306, 309  
Trans-New Guinea, 7, 198, 217, 219,  
228, 229, 234, 235, 248, 249,

*Language index*

- 255, 257, 261, 263, 272  
Tregami, 285, 314  
Tsakhur, 63, 71–73, 88  
Tsez, 38, 42  
Turkana, 26  
Turkic, 8, 280, 285, 314, 318  
Turkish, 85  
Tuu, 100, 140  
Tuwuli, 130  
T'apo, 147  
  
Ubangi, 115  
Uduk, 6, 36, 147–151, 153, 157, 160,  
162–167  
Upper Cross, 113  
Ushojo, 284–288, 291, 306, 309  
Uzbek, 285, 314  
  
Waigali, 285, 314  
Wakhi, 285, 313, 315  
Walman, 6, 171–173, 175, 177, 179, 181,  
182, 186–190, 192–195, 229,  
232, 238, 242, 246, 247, 264,  
269  
Warapu, 229, 232, 238, 239, 241, 242,  
246, 247, 251, 264, 269  
West Caucasian, 78  
West Papuan, 228–230, 237, 239  
Wogamusin, 268  
Wolof, 114, 115  
Wotapuri-Katarqalai, 282, 284, 286,  
288, 291, 309  
  
Yabus, 148  
Yava, 228  
Yidgha, 283, 312  
Yimas, 78, 229, 232, 237–239, 241,  
242, 245–247, 250, 251, 256,  
257, 264, 269, 270  
  
Yukaghir, 80  
Yukaghir, Kolyma, 80  
Yukaghir, Tundra, 79–81  
Yurok, 75, 76  
  
Zulu, 46  
  
!Xóõ, 165



# Did you like this book?

This book was brought to you for free

Please help us in providing free access to linguistic research worldwide. Visit <http://www.langsci-press.org/donate> to provide financial support or register as a community proofreader or typesetter at <http://www.langsci-press.org/register>.





# Grammatical gender and linguistic complexity

The many facets of grammatical gender remain one of the most fruitful areas of linguistic research, and pose fascinating questions about the origins and development of complexity in language. The present work is a two-volume collection of 13 chapters (plus an introductory chapter in each volume) on the topic of grammatical gender seen through the prism of linguistic complexity. The contributions discuss what counts as complex and/or simple in grammatical gender systems and whether the distribution of gender systems across the world's languages relates to the language ecology and social history of speech communities. Contributors demonstrate how the complexity of gender systems can be studied synchronically, both in individual languages and over large cross-linguistic samples, and diachronically, by exploring how gender systems change over time. In addition to three chapters on the theoretical foundations of gender complexity, volume one contains six chapters on grammatical gender and complexity in individual languages and language families of Africa, New Guinea, and South Asia.

This volume is complemented by volume II: *World-wide comparative studies*, which consists of three chapters providing diachronic and typological case studies, followed by a final chapter discussing old and new theoretical and empirical challenges in the study of the dynamics of gender complexity.

ISBN 978-3-96110-178-8



9 783961 101788