

Chapter 5

The dynamics of gender complexity

Bernhard Wälchli

Francesca Di Garbo

Stockholm University

In this chapter we view grammatical gender as a category type that emerges, evolves and disappears in languages as a result of diachronic processes and whose complexity grows and diminishes through time (§1–§2). Traditional approaches to grammatical gender focus on two properties that already presuppose a high degree of maturity of gender systems: noun classes and agreement. Here we conceive of gender rather as a category type with a semantic core of animacy and/or sex reflecting classes of referents, which have a propensity to turn into classes of noun lexemes. When growing and retracting, gender characteristically follows the animacy or individuation hierarchy. However, this hierarchical patterning breaks down when animacy leaks into the inanimate domain led astray by many different associative pathways, which is why lexical organization according to noun classes has to be invoked to maintain some sort of order (§3). Gender manifests itself in the form of marking on noun-associated words, often within the local domain of noun phrases. Here we put gender marking into the wider context of nominal morphology (non-lexical markers within the noun phrase), which often originate in independent use in headless noun phrases and are extended to headed noun phrases only in a subsequent development (§4). As more mature manifestations of gender get organized in the form of noun classes, they typically follow certain pathways of development that can be subsumed under the formula “From X to Y” (§5–§6). Agreement is fuzzy as its prototypical non-noun targets gradually develop by way of decategorialization from nouns, and controllers and targets are not always simple words, but can be complex (consist of syntactic formal groups) and controllers can be entirely contextual (§7). Gender should not be considered in isolation as it is – more often than not – parasitic on other grammatical category types, notably number, case, and person, with which it cumulates and which contribute to its high degree of complexity (§8). Number is particularly tightly intertwined with gender in pluralia tantum and other phenomena related to lexical plurality (§9). As gender is



organized in form of systems, its diachronic evolution cannot be captured in terms of individual diachronic processes. When gender systems evolve, there is virtually always co-evolution of connected events. Hence the study of system evolution is indispensable for understanding the complexity of gender (§10). However, the evolution of gender also displays characteristic areal and genealogical patterns and is sensitive to external factors of language ecology (§11).

Keywords: gender, complexity, animacy, historical linguistics, agreement, number, pluralia tantum, system emergence, areal linguistics, language ecology.

1 Introduction

This chapter has no ambition to provide a comprehensive survey of the very rich literature that exists on grammatical gender, for which we refer to Corbett (1991; 2006; 2014), Aikhenvald (2000; 2016), Kilarski (2013), Heine (1982), and Seifart (2010), to mention just a few. Furthermore, no attempts are made here to strictly delimit gender from classifiers; rather, grammatical gender is our focus of interest. Moreover, this chapter does not relate grammatical gender to gender studies. Having stated what this chapter is NOT about, let us now proceed to explain its focus of interest.

This chapter represents a DYNAMIC APPROACH to the understanding of grammatical gender (henceforth simply called *gender*). This means that we view gender as something that emerges, evolves and disappears in languages as a result of diachronic processes. Greenberg (1978) has been an important source of inspiration for the kind of diachronic and dynamic approach we propose here. In addition to the diachronic perspective, we are also interested in assessing the complexity of gender. While in many languages gender is complex, which is why Corbett (1991: 1) calls it “the most puzzling of the grammatical categories”, different degrees of gender complexity are attested in different languages. There are also languages with simpler kinds of gender.

In this chapter we are interested in why gender can grow quite complex in some languages and remain rather simple or turn simple again in other languages. Thus, even as far as complexity is concerned, we adopt a dynamic approach. We view gender as a MATURE PHENOMENON. According to Dahl (2004: 2), a mature phenomenon is a phenomenon that presupposes a non-trivial prehistory. Since we are also interested in how gender comes into being in the first place, we cannot define the object of study too narrowly, as otherwise there is a risk that we will miss much of the non-trivial prehistory. Our approach to linguistic complexity, in general as well as in the domain of gender, is outlined in §2. In the

following, we provide a roadmap for the topics discussed in the chapter and how they relate to the general purposes of this two-volume work.

At least since Hockett's (1958: 231) succinct definition – “Genders are classes of nouns reflected in the behavior of associated words”, adopted by Corbett (1991) – it has been common to define gender in terms of NOUN CLASSES and AGREEMENT. We argue here that noun classes and agreement are both mature phenomena.

The prototypical function of nouns is to express referents (Croft 2005: 438; Baker 2003). Hence, there are two different things that can be meant by noun classes: classes of noun lexemes and classes of referents. These are manifest in Dahl's (2000a: 107) notions LEXICAL GENDER, classes of noun lexemes, and referential gender, for which we use the name REFERENT-BASED GENDER suggested by Nichols (2019 [in Volume I]), classes of referents. Most approaches to gender take for granted that lexical gender is the primary object of interest, as, for instance, reflected in Corbett & Fedden's (2016: 9) Canonical Gender Principle: *in a canonical gender system, each noun has a single gender value*. It is not a priori clear why it should be useful for a language to partition noun lexemes into classes, but it is immediately understandable why speakers may be inclined to classify real world objects into classes. If we adopt a dynamic approach to gender, it is thus a reasonable assumption that referent-based gender is primary and that lexical gender is a later development that does not really have any clear purpose, but is somehow hard to avoid once words bearing gender markers are constantly associated with nouns and constantly collocate with nouns. The relationship of gender and reference is discussed in §3.

Many researchers agree that gender always has a SEMANTIC CORE: ANIMACY and/or SEX (Dahl 2000a: 101; Corbett 1991: 68; Luraghi 2011). However, somewhat strangely, this semantic core is usually not considered part of the definition of gender. The male-female sex distinction is clearly connected to animacy, as it is not applicable strictly semantically to inanimates. Animacy is thus crucial for the organization of gender. Animacy is a hierarchy rather than a simple dichotomy. Hierarchies are principles of organization that can considerably limit the complexity of a phenomenon. Hence, an important question for us to consider is how the ANIMACY HIERARCHY relates to the complexity of gender. As far as reference and lexicon are concerned, it makes more sense to organize referents according to a semantic core, and notably according to the animacy hierarchy, than noun lexemes. In languages where nouns carry grammatical markers, the DECLENSION CLASSES that structure grammatical allomorphs need not adhere to any semantic principle. This can be taken as evidence that classes of referents are crucial for the understanding of gender. The hierarchical patterning of gender is also discussed in §3.

Like other grammatical category types, gender is expressed by grammatical markers, viz. GENDER MARKERS. Unlike declension classes, these are not directly realized on nouns that condition the choice of class (Güldemann & Fiedler 2019 [in Volume I], use the term “deriflection”), but on NOUN-ASSOCIATED FORMS (adnominal modifiers, verbal argument indexes, or anaphoric pronouns, to mention just the most important ones). Many noun-associated forms are parts of the NP, so gender marking has to do with the wider question of what kind of non-lexical marking exists within noun phrases and how this marking emerges. There are languages that get along perfectly well without any NOMINAL MORPHOLOGY (non-lexical markers within the noun phrase). Nominal morphology is obviously a mature phenomenon. However, unlike its sub-phenomenon gender within noun phrases, nominal morphology need not necessarily distinguish classes. It can be the same marker all over, as in the English prop-word *one* for independent adjectives (adjectives without overt nominal head), as in *the big one*. Lehmann (1982), Moravcsik (1994), and others have emphasized the importance of independent noun-associated elements (such as free relative clauses, pronominal demonstratives and numerals) for the development of markers on attributive modifiers. From a developmental perspective on gender, it is important to put gender into the broader context of how nominal morphology emerges and spreads across various kinds of elements in the noun phrase. This is what we discuss in §4.

Conceiving of NOUN CLASSES in a dynamic perspective means to view them as phenomena undergoing change, which can be expressed by the formula “FROM X TO Y”. As already mentioned above, noun classes typically change from referent-based to predominantly lexical, and sometimes back to referent-based gender again (as in English). Several types of changes in noun classes have in common that there is an increase of complexity, notably the development of several types of gender assignment, the development from semantic to opaque assignment (gender assignment characterized by numerous exceptions), and the generalization of noun classes to all nouns. The dynamics of gender assignment and its evolution are what we focus on in §5 and §6.

The word *complex* is ambiguous. Most of the time we are talking about complexity in this chapter we mean by it (i) non-trivial in structure, so that an exhaustive description cannot be short. But “complex” can also mean (ii) consisting of several elements and (iii) consisting of different, but related phenomena. In discussing agreement, meanings (ii) and (iii) will be as important as meaning (i) and it is important to discuss how they relate to each other. Since Corbett’s (2006) influential monograph, AGREEMENT has often been conceived of as a morphosyntactic feature, which emphasizes the morphological realization on word-forms

and the syntactic nature of the link between controller and target. Given that there is not only intra-sentential, but also inter-sentential agreement, we hold that agreement is less uniform than commonly believed. It can be both syntactic and semantic. However, controllers and targets are not just morphological units, but often consist of several words (complex controllers and complex targets), and hence syntactic rather than morphological units. Controllers can be latent and are then neither morphological nor syntactic, but entirely contextual elements. It is often claimed that agreement always expresses coreferentiality, but coreference is actually only one of several specific relationships that may hold between controller and target. While controllers are typically nouns and targets noun-associated words, there are also nominal gender targets, and it can be shown that agreement often emerges step-by-step when nouns *DECATEGORIALIZE* (lose their nominal properties). To put agreement into a dynamic perspective means to recognize that agreement is not a uniform phenomenon, but rather a family of similar phenomena with complex diachronic relationships among them. We therefore suggest a broad definition of agreement, since a narrow definition is not easily compatible with a dynamic approach. This is the topic of §7.

Every definition of gender faces the problem that there is not just one, but several other grammatical category types that gender interacts with. Many researchers have recognized the close relationship to *CLASSIFIERS*, and it has even become common to view gender and classifiers as one set of phenomena, for which various cover terms have been proposed, such as nominal classification (Seifart 2010) and nomifiers (Haspelmath 2018). At least since Dixon (1982: 160), it has been common to argue that gender is characterized by a smallish number of classes (usually between two and ten, but sometimes up to twenty) and by obligatory grouping of all nouns into noun classes. A possible dynamic interpretation would be that gender is just a more advanced stage in the grammaticalization of nominal classification than classifiers (Passer 2016b). However, there is much reason to believe that many gender categories never went through a classifier stage (Nichols 1992: 142). While it is undeniable that some phenomena are intermediate between genders and classifiers, a major problem of the unified account is that gender does not entertain close relationships only to classifiers, but also to a range of other grammatical category types; for instance, with indexation (Croft 2003, Croft 2013; see §7.1) and with person name markers (markers indicating that an element is the name of a person). It has been repeatedly observed that a majority of languages with gender exhibit *CUMULATION* with *NUMBER*, and cumulation of gender and *CASE* and of gender and *PERSON* is also very common. This trend is so far-reaching that we think it is reasonable to include “cumulation with

number, case and/or person” into the definition of gender (notably since such cumulation is often lacking in classifiers). In fact, to the extent that it is known how gender systems evolve, cumulation with number or case often exists from the very beginning. One reason for this is that animacy is a typical conditioning factor for the choice of number and/or case (for instance, in differential object or other differential case marking). From a condition on number or case, animacy can further develop into a gender feature (a fully paradigmaticized grammatical category type expressed by systematic morphological marking) that still maintains cumulative exponence with the grammatical categories it originates from. This suggests that gender can be mature and hence complex from the very beginning and just appropriates the complexity of other mature grammatical categories it is connected with. Thus, when we say that gender is mature this does not necessarily entail that there is a non-trivial prehistory of gender, it can be a non-trivial prehistory of another grammatical category. Cumulation of gender with number, person, and case is discussed in §8.

Beyond the patterns of cumulative exponence that make gender closely interact with the encoding of number, case, and person, *PLURALIA TANTUM* nouns, that is, nouns that only exist in the plural, and other phenomena related to *LEXICAL PLURALITY*, whereby plural nouns form lexical classes, may also pose delimitation problems to the definition of gender as an independent grammatical category type. A common approach is to do away with this delimitation problem by saying that *pluralia tantum* cannot be a gender because their special behavior stems from them being lexically specified for number, which is a separate morphosyntactic category. This way of thinking derives from the assumption that gender and number are different morphosyntactic features. However, there is growing evidence that there are languages with two largely independent *CONCURRENT GENDER SYSTEMS* which cannot be subsumed under one gender feature (Fedden & Corbett 2017; Corbett et al. 2017; Svärd 2019 [in Volume I]; Liljegren 2019 [in Volume I]). If there is not just one gender feature, why then should we assume a priori that gender and number features must always be neatly distinct? There is evidence from a dynamic perspective that *pluralia tantum* can develop into gender classes diachronically, which is an argument for a close relationship between gender and lexical plurality (Dryer 2019, Olsson 2019, both in Volume I of this work). The relationship between gender and *pluralia tantum* is discussed in §9.

Gender is often called a “SYSTEM”, but few approaches are explicit in what this label implies. A system is minimally an opposition between at least two markers, but mature gender systems are more complex than that. They are highly orga-

nized language-specific complexes with both paradigmatic and syntagmatic components that play an important role in the architecture of grammar. Although systems can exhibit considerable complexity, there is reason to believe that they are also mechanisms to keep complexity within manageable limits. For the dynamic approach it is important to view systems as phenomena that emerge and evolve. Hence, rise, expansion, reduction and loss of gender must be viewed as processes of system evolution. This is the topic of §10.

However, the structure of gender does not only have language-internal implications. Gender exhibits specific GENEALOGICAL AND AREAL PATTERNS. It has repeatedly been observed that gender is quite stable diachronically, but gender seems to be more stable in a language if the contact languages also have gender systems of the same kind. A further question is whether there are any external factors in the ECOLOGY OF LANGUAGES that condition whether languages have gender and what kind of gender systems. This and related questions are addressed in §11.

Having provided a roadmap for the main topics discussed in this chapter, we are now in a position to propose a tentative definition of gender that takes the dynamic approach into account.

Gender is a grammatical category type with a semantic core of animacy and/or sex reflecting classes of referents, which have a propensity to turn into classes of noun lexemes. It is overtly marked on noun-associated forms. It typically exhibits cumulative exponence with number, case, and/or person. Gender is organized in the form of systems.

The building blocks of this dynamic definition of gender are discussed in the remainder of this chapter and based on the following outline. §2 considers the relationship between gender and complexity. §3 explores the relationship of gender with reference and animacy. §4 discusses gender in the broader context of nominal morphology. §5 and §6 deal with noun classification and gender assignment. §7 reconsiders the notion of gender agreement. §8 investigates the relationship between gender, number, case and person while §9 focuses on pluralia tantum. §10 explores the extent to which gender is subject to system evolution. §11 addresses gender in its genealogical and areal context and discusses the relevance of external factors in the ecology of languages. §12 summarizes the results and concludes the chapter.

Since *gender* is not the only term to keep track of, we have compiled an appendix with short definitions of terms at the end of this chapter. All definitions have the perspective of gender and/or complexity and their primary purpose is

to facilitate the understanding of this chapter rather than being universally applicable in linguistics.

2 Complexity and gender

In §2.1 we provide an overview of current approaches to the notion of linguistic complexity. §2.2 then discusses the relationship between complexity and gender, as well as the existing metrics of gender complexity.

2.1 Understanding and measuring complexity

Over the last couple of decades, the debate on linguistic complexity has focused primarily on three overarching topics:

- (i) what counts as linguistic complexity,
- (ii) what to measure when quantifying complexity,
- (iii) and what relevance this has for understanding languages overall.

These topics, and their relevance to the understanding of grammatical gender, are tackled in the two volumes of this work, and, more specifically, in the chapters by Audring, Nichols and Sinnemäki. Audring (2019 [in Volume I]) provides a theoretical account of gender system complexity by comparing the notion of linguistic complexity with canonicity and difficulty. Nichols (2019 [in Volume I]) tests and falsifies the hypothesis that languages with gender are more complex overall. Sinnemäki (2019 [this volume]) investigates whether there is a complexity trade-off between the distribution of gender systems and that of numeral classifiers across the languages of the world. A fourth contribution, Di Garbo & Miestamo (2019 [this volume]), approaches gender system complexity from a diachronic perspective by investigating disappearing and/or emerging patterns of gender agreement and their complexity features.

Starting with the first topic – what counts as linguistic complexity – all four contributions define complexity in *ABSOLUTE* terms, that is as an objective property of grammatical domains rather than as a subjective feature of language use (what is also known as *relative complexity*). This issue has been extensively debated in the literature. While some influential cross-linguistic studies in the field (Kusters 2003; 2008) deal with complexity as a measure of difficulty in language learning and use, the dominant approach in the functionally-oriented literature

has been that linguistic complexity is best viewed as a property of language systems, rather than as a measure of ease of acquisition and use. This is essentially because we do not yet have a full account of language processing difficulties in different domains of grammar and across different modes of language acquisition. Important contributions in establishing the roadmap for such an approach to the theoretical and empirical study of linguistic complexity are the two volumes edited by Miestamo et al. (2008) and Sampson et al. (2009).

The second issue that has been central in the debate on language complexity is what to measure when quantifying complexity. Nichols (2019 [in Volume I]) sees two main answers to this question: (i) *INVENTORY* or (*COMPOSITIONAL*) *COMPLEXITY*, that is, the number of distinctions in a grammatical system (e.g., the number of tones, tenses, genders), and (ii) *DESCRIPTIVE COMPLEXITY* (or Kolmogorov complexity), defined as the information required to describe a system (the longer the description, the more complex the system). While Nichols sees inventory and descriptive complexity as independent of one another and argues that descriptive complexity “is a better measure” that “captures well the non-transparency relevant to learnability and prone to be shaped by sociolinguistics”, an integrated approach is proposed by Miestamo (2008), and followed by Audring (2019 [in Volume I]) and Di Garbo & Miestamo (2019 [this volume]). Under this approach, linguistic complexity is defined in terms of overall description length, which can be measured on the basis of two principles, the Principle of Fewer Distinctions and the Principle of One-Meaning–One-Form.

The *PRINCIPLE OF FEWER DISTINCTIONS* is a measure of inventory complexity and states that the fewer distinctions are made within a grammatical domain the less complex the domain (the fewer the tones, tense or gender distinctions, the less complex the tone, tense or gender system overall). The *PRINCIPLE OF ONE-MEANING–ONE-FORM* is a measure of transparency whereby the less complex grammatical phenomenon is one where there is a one-to-one correspondence between meaning and form.

Under the Principle of One-Meaning–One-Form, cumulative morphemes (simultaneously expressing more than one grammatical meaning) or multiple exponents (where one grammatical meaning is distributed over several morphemes) are more complex than morphemes that are only associated with one grammatical meaning. Working specifically on the measurability of gender system complexity, Di Garbo (2014; 2016) expands this approach by proposing a third complexity principle, the *PRINCIPLE OF INDEPENDENCE*, which targets interactions between grammatical domains and their effect on the overall complexity of individual domains. Under the Principle of Independence, a marker that cumulates

the encoding of gender and number distinctions features higher gender complexity than a non-cumulative marker, because the marking of gender distinctions is dependent on the number value of nouns. In the gender complexity measure proposed by Audring (2019 [in Volume I]), the Principle of Fewer Distinctions and the Principle of One-Meaning–One-Form are referred to as Principle of Economy and Principle of Transparency, while the same label as Di Garbo (2014; 2016) is kept for the Principle of Independence.

The third and final issue that has been frequently addressed in the literature on linguistic complexity is the relevance that complexity measures may have to understanding languages overall. Is it at all possible to design complexity metrics that allow us to estimate whether one language is generally more complex than another? And, provided that this is the case, how can such metrics be used? McWhorter (2001) proposes to measure overall grammatical complexity on the basis of a pool of features ranging from phonology to syntax. However, the features suggested by McWhorter (2001) aim to capture the peculiarities of one specific language profile, the creole profile, and to demonstrate that creole languages are overall less complex than non-creoles. The question thus remains whether the ambition to measure linguistic complexity overall is a feasible, and even meaningful, enterprise even beyond the creole/non-creole dichotomy. This question is approached in work by Miestamo (2008), Nichols (2009), and Sinnemäki (2014b), who argue that measures of global linguistic complexity are both theoretically and empirically unfeasible. Even assuming that the daunting task of formulating an exhaustive inventory of complexity features that are truly representative of overall grammatical complexity could be accomplished, it is still hard to establish empirically how each of these features contributes to overall complexity in comparison to others. For instance, it would be impossible to truly establish whether the presence of grammatical gender implies higher complexity than the presence of, say, grammaticalized tone distinctions, or the other way round, both within and across languages. Miestamo (2008) refers to this as the problem of comparability, and argues that one way to overcome this problem is to restrict the quantitative and qualitative typological study of linguistic complexity to individual grammatical domains, and eventually compare domain-specific data with each other in search of potential complexity trade-offs between individual grammatical domains and their functional explanations. All contributions to this two-volume work approach the complexity of gender systems in the spirit of this suggestion, and even those chapters that explicitly focus on comparisons and relationships between the complexity of gender and other domains of grammar bring support to the idea that domain-specificity is a key to understanding the distribution of linguistic complexity within and across languages.

The chapter by Nichols (2019 [in Volume I]) shows that testing whether the presence of grammatical gender makes languages more complex overall produces negative results. Sinnemäki (2019 [this volume]) demonstrates that comparing gender and classifier systems in terms of complexity distributions may be useful to unravel functional trade-offs in the domain of nominal classification, whereby the presence of grammatical gender in a language disfavors the occurrence of numeral classifiers and vice versa.

2.2 Gender complexity metrics and the principles behind them

Gender complexity metrics have been proposed by Audring (2014; 2017) and Di Garbo (2014; 2016). In their approaches, gender is considered to be a grammatical domain of its own, and its complexity is assessed on the basis of the three principles introduced in §2.1: (i) Economy, or the Principle of Fewer Distinctions, (ii) Transparency, or the Principle of One-Meaning–One-Form and (iii) (the Principle of) Independence. See Audring (2019 [in Volume I]) and Di Garbo & Miestamo (2019 [this volume]) for a more detailed treatment of the three principles.

Di Garbo's (2014; 2016) metric is an index of six features, each ranging between zero and one, that is applied to a dataset of 84 African languages with gender. The features are Number of gender values (GV), Nature of assignment rules (AR), Number of indexing targets (IND) (all three Fewer Distinctions), Cumulative exponence of gender and number (CUM) (One-Meaning–One-Form and Independence), Manipulation of gender assignment triggered by number/countability (M1), and Manipulation of gender assignment triggered by size (M2) (both Independence). The first three features of the metric are based on the proposal by Audring (2014). Features CUM, M1, and M2 are meant to measure the impact that interactions of gender and number, and gender and evaluative morphology, have on the overall complexity of gender. The features by Di Garbo (2014; 2016) are designed such that a simpler gender system can always take the value zero (only two genders, only semantic assignment, only one indexing target, non-cumulative exponence and no manipulation of gender). However, all languages in Di Garbo's sample have higher total values than 0.0, and many have 1.0, which can be interpreted such that gender tends to be complex at least in African languages. The metric has been applied by Liljegren (2019 [in Volume I]) to the languages of the Greater Hindu Kush area. He identifies two languages with value 0.0, Khowar and Kalasha (together making up the Chitral subgroup), both of which have developed animacy-based gender distinctions quite recently. All other gender languages in Liljegren's sample have Medium or High gender complexity.

Based on the same three principles, Audring (2017) develops a metric consisting of 23 features, which all can take the values *simple* and *complex*. The metric is illustrated only for one language, German (Indo-European, Germanic), whose gender system turns out to score less than expected (only 9 of 23 features complex). This is most likely due to the fact that many simple gender features have been overlooked in the literature, while the metrics proposed by Audring allows to capture them. This metric is further elaborated upon by Audring (2019 [in Volume I]) in the context of a broader discussion of the relationships between complexity, canonicity, and difficulty.

3 Referent-based gender and the limited hierarchical patterning of gender

3.1 Introduction

In this section we are going to argue that REFERENT-BASED GENDER (classes of referents) is more basic from a developmental perspective than LEXICAL GENDER (classes of noun lexemes). Referents are typically classified in terms of animacy, and animacy is organized in form of a hierarchy. Nichols (2019 [in Volume I]) suggests that hierarchical patterning is a decomplexifying mechanism. The question thus arises as to why hierarchical patterning in gender does not limit complexity. In this section we will argue that hierarchical patterning in gender is rooted in referent-based gender and that gender typically originates as referent-based gender in the top segment of the animacy or individuation hierarchy (§3.2). However, as referent-based gender travels down the animacy hierarchy, there are two things that happen that render it less transparent. First, some or several animal gender values (animate, or, masculine and/or feminine) are expanded to inanimate objects by means of various factors often of a metonymic or metaphoric character, such as agentivity, discourse salience, uniqueness, power, purview, and possession (discussed in §3.3 and §3.4). These cannot be neatly arranged on a single scale and hence hierarchical patterning fails to apply to them. Second, if gender has travelled down the animacy hierarchy, it gets increasingly more associated with nouns and is aligned with the conceptual structure of nouns, which means that it turns into lexical gender (§3.5). Noun lexemes, however, are not subject to hierarchical patterning in the same way as referents, and hence there is no hierarchical mechanism that can efficiently limit the complexity of lexical gender, even though the semantic core originating in referent-based gender is maintained. In order to prevent complete disorder, gender must thus resort

to lexical patterning, instead of hierarchical patterning, and lexical patterning has a much higher degree of freedom than hierarchical patterning, which entails that complexity is less effectively limited in lexical gender. However, when lexical gender develops, referential-based gender does not disappear, but interacts with it, and lexical and referential-based gender are sometimes so similar that they are difficult to distinguish (§3.6). Both lexical and referent-based gender allow for reconceptualization of referents, which is why gender is not suitable for reference tracking (§3.7).

3.2 The animacy/individuation hierarchy

Dahl (2000a) follows Aksenov (1984) and Corbett (1991) in claiming that all gender systems have a semantic core. He uses the animacy hierarchy in (1)

- (1) Animacy hierarchy
 human > higher animals > lower animals > inanimate

to further specify that core. Above some cutoff point on the animacy hierarchy, gender is semantically assigned; below the cutoff point, gender is non-semantic (formal or arbitrary). If the animate pole is further subdivided, the major criterion is sex. From the point of view of complexity, this means that gender tends to be simple on the animate pole of the hierarchy and complex on the inanimate pole of the hierarchy, even though often the same gender values are used both above and below the cutoff point: “inanimate nouns are quite often assigned to genders whose semantically determined core consists of animates” (Dahl 2000a: 102–103) and “gender distinctions often cut through the animal kingdom” (Dahl 2000a: 100). A neat example is Walman (Nuclear Torricelli, West Palai; see Dryer 2019 [in Volume I]), where nouns denoting humans and some larger animals are either masculine or feminine, depending on the sex of the referent, whereas nouns denoting most animals, especially non-mammals, appear to have relatively arbitrary gender, but are assigned to the same two genders masculine and feminine.

Various forms of the animacy hierarchy can be found in the literature. Croft (2003: 130) uses the so-called extended animacy hierarchy (2), which, according to him, combines three distinct, but related hierarchies: person, referentiality, and animacy (3). Here and elsewhere we replace “proper names” by *person names*, since names of animals, things, and places are usually disregarded in discussions of animacy.

- (2) Extended animacy hierarchy (Croft 2003: 130, following Dixon 1979: 85 and Silverstein 1976):
 first/second person pronouns > third person pronoun > person names > human common nouns > nonhuman animate common nouns > inanimate common nouns
- (3) Component hierarchies of the extended animacy hierarchy
 Person: first, second > third (proximate > obviative)
 Referentiality: pronoun > person name > common noun
 Animacy: human > higher animals > lower animals > inanimate

As pointed out by Croft (2003: 166), different hierarchies often interact. He illustrates this with examples from Eastern Panjabi¹ (Indo-European, Indo-Aryan) for differential object marking, which combines the factors animacy and referentiality. Eastern Panjabi objects are overtly coded with (-) *nū* unless the object is both inanimate and non-definite (specific or non-specific). This suggests that the component hierarchy “Referentiality” actually falls into two different sub-hierarchies: Part-of-speech (pronoun > person name > common noun) and Definiteness (unique > definite > specific/referential > non-specific/non-referential).

Siemund (2008) surveys pronominal gender in varieties of English and other languages. He comes to the conclusion that pronominal gender in English crucially depends on the degree of individuation of the entries referred to. He adduces Sasse’s Individuation Hierarchy (1) to account for this, a further variant of the animacy hierarchy.

person names	humans	animals	inanimate tangible objects	abstracts	mass nouns
humans		non-humans			
animates			inanimates		
count nouns				mass nouns	

Figure 1: Individuation hierarchy according to Sasse (1993: 659), as adapted by Karatsareas (2014: 90)

The individuation hierarchy is more elaborate than Croft’s extended animacy hierarchy in that it contains two further sub-hierarchies: countability (count noun > mass noun) and concreteness (concrete/tangible > abstract).

¹Here and elsewhere in this chapter, we use mostly the language names in Glottolog in Hammarström et al. (2018). Thus, here “Punjabi” is replaced by “Eastern Panjabi”. If Glottolog names are not used, these are given in brackets.

Individuation is relevant in varieties of English in various ways. Diachronically, Siemund & Dolberg (2011: 527) show that “gender change appears to have started with nouns ranking low in terms of individuation” in English in the transformation of a German-like complex gender system to a pronominal gender system. In West Somerset English there is a mass/count distinction (*the bread – it, the loaf – he*). For further examples from other languages, see Siemund (2008: 175–217).

3.3 The animacy hierarchy does not structure the connections between inanimate referents and animacy

The extended animacy hierarchy does not provide any guidelines for the domain of inanimates and the individuation hierarchy offers only three very general groupings: tangible objects, abstracts and mass nouns. However, there are various semantic connections that can link inanimate referents with animates, notably the following: agentivity, salience, purview, uniqueness, power, and possession. Possession will be addressed in §3.4, the others in this section. In these semantic domains, there are often metaphorical or metonymic connections between inanimate referents and animacy. These cannot easily be arranged on a single scale of animacy and individuation. But they all provide pathways for expansion from animate genders to the inanimate domain.

- (i) *Agentivity*. Agents are usually conceived of as animate. Inanimate referents can leak into animate genders when they are construed as agents. In Zande (Atlantic-Congo, Ubangi), nouns referring to inanimate things usually control inanimate gender, but can take animate gender when animacy is imputed on them, as in (4):

- (4) Zande (Atlantic-Congo, Ubangi; Gore 1926: 32)
Ime ki sa ti-ru (/ti-e) ni kure.
 water and.then turn REFL-ANIM (/REFL-INAN) with blood.
 ‘And the water turned itself into blood.’

In various languages with gender, nouns for natural phenomena are treated as animates. In his description of gender in Walman, Dryer (2019 [in Volume I]) mentions that the quasi-animate natural phenomena *onyul* ‘earthquake’, *knum* ‘whirlpool, riptide’, *snar* ‘moon’, and *nganu* ‘sun’ are masculine (unlike all other inanimates, which are feminine). *Nganu* does not only mean ‘sun’, but also ‘day’, and is masculine in both meanings,

hereby demonstrating that the gender of this noun, even though motivated by animacy and probably originating from referent-based gender, has become lexical gender.

Agentivity can come in different forms. It can be more syntactic, as in (4) where an inanimate referent is construed as an actor, or it can be more derivational, when an inanimate referent is construed as an agentive noun. Mopán Maya has masculine and feminine person name markers extended to common nouns, and one of their major functions is to form analytic agentive nouns: *ix p'o'* [GM.F wash] ‘washerwoman’ (“Ms wash”), *aj jook'* [GM.M fish] ‘fisherman’ (“Mr fish”) (Contini-Morava & Danziger 2018: 140). Gender markers can be used to suggest unexpected agentivity of inanimate objects as in (5).

- (5) Mopán Maya (Mayan, Yucatecan; Contini-Morava & Danziger 2018: 141)

<i>Ox-tuul-oo'</i>		<i>aj</i>	<i>kuch-b'äk'</i>	<i>a</i>
three-NUM.CLF.ANIM-3UNDERGOER.PL		GM.M	carry-meat	ART
<i>xoolte'</i>	<i>leek-oo'</i>		<i>a</i>	<i>b'e'</i>
walking_stick	3.EMPH-3UNDERGOER.PL		ART	DEM.4

‘Those aforementioned walking sticks became three (living) meat-carriermen.’

- (ii) *Salience*. Pawley (2004) shows that “animated pronouns” are a prominent feature in Tasmanian Vernacular English. That there is some degree of animation can also be seen from examples where other animate words besides pronouns, such as *fella*, are used, as in (6). “The entity must be referential (specific or definite). Other factors include its importance in the discourse (as a main topic, background element, etc.), its sequential position in the discourse, and its inherent salience” (Pawley 2004: 114). For portable goods other than vehicles, *he* expresses an attitude of detachment and *she* (emotional) attachment. Plants, animals, and male genitals are *he*, everything else is *she*.

- (6) Tasmanian Vernacular English (Pawley 2004: 126): attitude of detachment
 [Salesman is showing carpets to two customers] *That fella he's a poly, he's two fifty.*

The classical study on animated pronouns in English is Mathiot & Roberts (1978), who observed similar patterns in spoken English in two parts of the USA, Los Angeles County and Buffalo NY. However, for portable goods there is a polarity effect: men use predominantly *she* and women *he* (Pawley 2004: 134).

- (iii) *Purview*. The notion of purview has been introduced by Gerdts (2013) for Halkomelem (Salishan, Central Salish). Halkomelem has a sex-based semantic gender system marked on determiners and demonstratives, where female singular humans take feminine gender and all other nouns masculine gender. However, feminine optionally appears on hundreds of inanimate nouns when they are in the feminine PURVIEW (Gerdts 2013). This includes objects that belong to or relate to a female, are perceived as being feminine in size, shape, or function, or are spoken about by a female.

In Comaltepec Chinantec (Otomanguan, Chinantecan), gender “can be rhetorically upgraded to express a closer association than the normal gender assignment would indicate” (Anderson 1989: 57). In (7), the word for ‘paper’, which is usually inanimate, is animate in order to mark a more intimate status as a product “of someone’s personal labor and attention” (Anderson 1989: 57).

- (7) Comaltepec Chinantec (Otomanguan, Chinantecan; Anderson 1989: 57)
- mi*^{LM-r} *hmi*^L*giú:n*^{L-b} *hiu*^{LH} *ma*^H*ht*^L
 request.3PL-3 many.ANIM-AFFIRMATION DIM paper
 ‘S/he asks for many papers.’

- (iv) *Uniqueness*. In the Irish-Canadian author Emma Donoghue’s (2010) novel *Room*, a mother and son are captured in a backyard shed that the boy never leaves until they manage to escape when the boy is five. In the boy’s language things in *Room* with unique reference are *he* and *she*. In *Room* English, feminine and masculine gender are inseparably tied to uniqueness and referentiality. In *Room*, *Blanket* is feminine: *we put Blanket over [TV] and just listen through the gray of her* (p. 11). However, things outside *Room* are all different and not unique: *it’s not fleecy gray like Blanket, it’s rougher* (p. 166), *Officer Oh tries to put the blanket over my head, I push it off* (p. 177). It may be argued that this example is artificial, but it is still a doculect of English, and the example shows that a particular use of referent-based gen-

der may be contextual and need not necessarily apply to a whole language as a system.

In narrative discourse, inanimates are often personified as unique referents. One of the arguments adduced by Leeding (1989: 232) that gender assignment in Anindilyakwa (Gunwinyguan) is semantically motivated is that masculine and feminine words often are connected in traditional Dreamtime stories as *dramatis personae*, e.g. *yi-ningwimwapwalhpwa akwa thi-wirrawilya* ‘M-Bat and F-Rainbow’. This suggests that they at least in some of their typical uses are conceived of as unique, which may have favored the extension of masculine and feminine to inanimates.

- (v) *Power*. Straus & Brightman (1982) have argued that the seemingly arbitrary distribution of animate and inanimate gender for inanimate referents in Algonquian languages is motivated by power. Animate nouns are all in some sense “powerful” (Straus & Brightman 1982: 135). In Cheyenne (Algonquian, Cheyenne), some body parts (finger/toe, thumb, fingernail, claw, eyebrow, knee, kidney, and brain) are animate, but not when indicated on a drawing or discussed as abstractions, and there is a good story for each of them why exactly these are powerful (Straus & Brightman 1982: 128–130). Cheyenne *mo?eško* ‘finger, toe’, for instance, is animate because fingers are used symbolically as weapons and as channels of power in cursing, but there is also an inanimate noun *mo?eško* ‘ring’. Power has cultural implications. However, whether ascribed to language or culture, the classification of nouns is complex and on some level due to convention. “For example, nouns labeling mechanical items introduced by Whites are largely inanimate in Ojibwa while they are often animate in Menomini” [Menominee] (Straus & Brightman 1982: 133). This suggests that referent-based animate gender of inanimates, originally motivated by the factor power, has largely turned into lexical gender in Algonquian languages.

3.4 Inherited gender

A further semantic connection between animate and non-animate referents is possession, which is even more difficult to include in the animacy hierarchy. In *INHERITED GENDER*, surveyed by Evans (1994) for Australian languages, the gender of a noun or NP is determined by the gender of its possessor. Inherited gender is usually referent-based rather than lexical gender, and this is stated explicitly by Olsson (2017: 186) for Coastal Marind (Anim, Marindic). In Coastal Marind,

a few nouns, including *igih* ‘name’ in (8), inherit the gender of the referent to which they are attached.

- (8) Coastal Marind (Anim, Marindic; Olsson 2017: 187)
igih ta/tu/ta/ti ka-ha-b
 name what:I/what:II/what:III/what:IV PRS.NEUT-INT-ACT[3SG.A]
 ‘What is his/her/its name?’ more literally: ‘What is the he-name,
 she-name, it-name?’

Note that the possessive pronoun in the English translation is misleading. The interrogative pronoun simply takes the gender that the speaker assumes to be the class of the referent (male name for gender I, female or nouns denoting animals for class II, thing of a noun in class III or IV for class III and IV; Olsson 2017: 187–188). For inherited gender in New Guinea, see also Fedden (2011: 177).

In Halkomelem, inherited gender is part of the extension of feminine to inanimates by purview. In example (9), the instrument *šaptən* ‘knife’ can be feminine if possessed by a female, but must be masculine when possessed by a male:

- (9) Halkomelem (Salishan, Central Salish; Gerdts 2013)
Ni’ ʔancə kʷθə/ɬə šaptən-s θey qemiʔ?
 AUX where DET/DET.F knife-3.POSS DEM.F girl
 ‘Where is that girl’s knife?’

In North America, inherited gender is also attested in Tunica (isolate), where body parts inherit gender (Swanton 1921: 23).

It is important to emphasize that inherited gender is not always referent-based gender. In Jarawara (Arawan, Madi) inalienable possession, the gender of the NP is determined by the gender of the possessor. The Arawan languages have complex lexical gender assignment. It is thus not surprising that even inherited gender in Jarawara is more complex than in Halkomelem, where there is virtually no lexical gender anywhere in the language. In Jarawara, it is the lexical gender of the possessor that is inherited, not the gender of the referent. Pronouns are feminine irrespective of referent-based gender (Dixon 2000: 489).² Hence, the NP in (10) is feminine and triggers feminine agreement on the predicate, whatever the sex of the referents. As we will see in §7.3, this is a kind of gender resolution. However, some inalienable nouns also have derivational gender suffixes, whose gender is determined in a different and rather complex way, but also by the lexical gender of the possessor. A first person inclusive possessor, as in (10), always

²Jarawara is not the only language where pronouns all trigger a specific gender. In Uduk (Koman; Killian 2019 [in Volume I]) pronouns are always in class 1.

triggers masculine derivational gender. This is an instance of a nominal target (see §7.6), where a derivational affix of a noun can be an agreement target.

- (10) Jarawara (Arawan, Madi; Dixon 2000: 490)
Ee man-o koma-ke.
 1PL.INCL.INALIEN arm-DERIV.M be.sore-DECL.F
 ‘Our (inclusive) arms are sore.’

In Australian languages it is common to mix inherited gender and intrinsic lexical gender. In Mawng (Iwaidjan Proper), some nouns for body parts, such as *ngijalk* ‘body’, always have inherited gender, whereas *ngaralk* ‘tongue’ (class IV), *murlu* ‘nose’ (class III) and *algij* ‘liver’ (class V) can have lexical gender (Capell & Hinch 1970; Evans 1994: 5). In Tiwi (isolate), body part nouns take the gender of their possessor, except for genitals that take the gender of the opposite sex (Evans 1994: 2). The opposite choice for genitals can be explained by purview. Genitals relate to the other sex. Opposite choice is also attested for the Amwi variety of War-Jaintia (Austroasiatic, Khasian; Weidert 1975): *?u kde* ‘DET.M vagina’, *kə khle* ‘DET.F penis’. However, inverted inherited gender, i.e. gender opposite to that of the referent of the possessor, is generalized in Amwi War. Body parts, tools and household items take the gender opposite to the person they are associated with. Inverted inherited gender in Amwi War exhibits the same fluidity as non-inverted inherited gender in Halkomelem, and this suggests that we have to deal with referent-based gender rather than with lexical gender here: *?u khlia kə* [DET.M head 3.SG.F] ‘her head’ (personal pronouns preposed to nouns are gender markers of that noun, possessors are postposed to their heads in NPs), *kə klia-w* [DET.F head-3.SG.M] ‘his head’. Only tools and clothes only associated with one sex are not fluid, in the same way as genitals: *kə cin* ‘jeans (only for men)’.

Inherited gender and gender by purview is a kind of associated gender. The most famous case of associated gender in the literature is Dyirbal. Corbett (1991: 16) uses the term concept association for the well-known Dyirbal examples where ‘fishing line’ and ‘fish spear’ are gender I (animate) because of their association with ‘fish’ (but see Plaster & Polinsky 2007 for an alternative explanation). In the light of the many examples of association by referent surveyed in this section it seems to us that the term “concept association” is problematic. Association in gender is mainly association with referents and not association with concepts. Of course, as in other cases where referent-based gender turns into lexical gender, association of referents can eventually turn into association of concepts.

Not only are the semantic connections discussed in §3.3 and this section often metonymical or metaphorical in character, some of them also provide pathways

for how cultural beliefs can make their way into language structure. This holds in particular for purview and power. If we conceive of gender as referent-based originally, it does not necessarily express cultural beliefs from its very origin. There are many languages with semantic gender assignment where there is no associated gender of the kind that is attested in Dyrbal and Algonquian languages. There is no reason to believe that communities speaking languages without associated gender are poorer in their cultural beliefs. It is thus possible to view the “culturalization” of gender as a trait of maturity. Languages with many cultural properties embedded in grammatical gender presuppose gender categories with non-trivial prehistories.

3.5 Lexical gender originates from referent-based gender

In §3.2 and §3.3 we have considered cases where referent-based animate gender leaks to inanimate referents. In this section we will now consider instances of referent-based gender marking that have further developed into lexical gender as gender has travelled down the animacy hierarchy from personal pronouns and person names to NPs headed by common nouns.

Russian and other Slavic languages have developed a lexical animacy distinction in addition to the three-way masculine-feminine-neuter lexical gender system inherited from Indo-European. Slavic animacy subgenera originate from differential object marking. Due to sound change, nominative and accusative singular came to be morphologically indistinguishable in the major masculine declension class, which is why forms of the genitive singular started being used in object function (Meillet 1897; Huntley 1980: 206), and the genitive form was then also used in non-object function following prepositions. In Old Russian of the 13th and 14th centuries, genitive singular forms had generally replaced accusative forms for personal pronouns and person names (Dietze 1973: 263). According to Dietze (1973: 265), socioeconomically subordinate and dependent persons, such as children, servants, slaves, and messengers, go with the inanimate category; *mužb* ‘man’ is animate in the meaning ‘husband’ but inanimate as the subjects of a prince, and *vinogradb* ‘vineyard’ is animate when used metaphorically for the world populated with people, as in (11). Hence, animacy gender on masculine singular common nouns was referent-based in Old Russian.

- (11) Old Russian (Indo-European, Slavic; Dietze 1973: 267)

<i>gospodi</i>	<i>bože...</i>	<i>posěti</i>	<i>svoego</i>
lord.VOC.SG	god.VOC.SG...	visit(PFV).IMP.2SG	OWN.GEN.SG.M

vinograda

vineyard(M/ANIM).GEN.SG

‘O Lord God...visit your vineyard’

Animate forms with animals start appearing in the 16th century, and in the 17th century, animate forms for animals were generalized (Dietze 1973: 270). (12) from Modern Russian illustrates that animate gender has become lexical. Even though the Modern Russian animacy distinction is clearly semantically motivated, it is entrenched in the lexicon and some modern dictionaries now indicate whether a noun is animate or inanimate. In (12) *konkurent* ‘competitor’ is animate (takes genitive singular in object function) and *Uzbekistan* is inanimate (takes nominative singular in object function), although they both have the same referent.

- (12) Russian (Indo-European, Slavic): lexically entrenched animacy distinction

Kazaxstan-Ø rassmatrivaet Uzbekistan-Ø kak

K.(INAN/M)-NOM.SG view(IPFV).PRS.3SG U.(INAN/M)-NOM.SG how/as

konkurent-a.

competitor(ANIM/M)-GEN.SG

‘Kazakhstan views Uzbekistan as a competitor.’

Gender marked on NP-markers may develop from person name markers. Person name markers have a tendency to be expanded. Varieties of Catalan have the person name markers masculine *en* (< *don* < Latin *dominus*) and feminine *na* (< *dona* < Latin *domina*). In Balearic Catalan, these markers can be expanded to names of animals (*en Pluto* for a male dog), and to folk names of clouds and celestial bodies: *en Catalí* ‘Venus at dawn’ (Caro Reina 2018: 195–197). This is arguably not lexical gender, since common nouns are not involved, but referent-based gender having traveled down the animacy hierarchy. Person name markers are very common in Austronesian languages, where they are sometimes extended to some older kinship terms, such as ‘father’ and ‘mother’, which are often unique, or titles.³ Like many other Austronesian languages, Tagalog (Central Philippine) makes a distinction between noun phrase markers for common nouns (topic *ang*,

³It should be noted here that uniqueness does not have the same effect in gender as in definiteness, where it has also been claimed to play an important role (Russell 1905; Lyons 1999). Definite articles are used in the first place with concepts that are not unique out of context, but which happen to be unique in a particular situation. Person name markers, however, express uniqueness on items that are unique in any context and are extended first of all to expressions that are typically construed as unique.

non-topic *nang*, oblique *sa*) and for person names (topic *si*, non-topic *ni*, oblique *kay*). Person name markers can also be used with older kinship terms (*ate* ‘eldest sister’, *kuya* ‘eldest brother’, *ina/nanay* ‘mother’, *ama/tatay* ‘father’ and some others); the difference is that this use is optional. They can be used with nouns designating occupations when expressing titles: *si Abogado Cruz* ‘Lawyer Cruz’ (Schachter & Otanes 1972: 94). In several Oceanic languages, person name markers have been extended to some common nouns and turned into lexical gender. In Nakanai, the person name marker is used with about 70% of the names of species of fish, birds and insects and a majority of loanwords (Johnston 1980: 166–167). For Teop, see Svård (2019 [in Volume I]) and the references given there; for Owa and Kahua, see §6.4; and for Austronesian in general, see Handschuh (2018).

In the Mek languages in New Guinea it can be observed how lexical gender can develop from a referent-based gender uniqueness vs. non-uniqueness distinction, where uniqueness looks very much like an extended person name marking. The more conservative Mek languages Una and Eipo [=Eipomek] have a uniqueness distinction, and Nalca, which is more progressive and closely related to Eipo, has developed a gender system with four lexical gender classes with rather simple gender assignment principles (Wälchli 2018).

Una has only an opposition between *bi-* unique and *a-* non-unique. *Bi-* is mainly used with person names and kinship terms older than ego – thus the cutoff point on the animacy hierarchy is between older and younger kin – but also with highly individuated non-kinship human nouns and sometimes even animals and things. Thus, in the Una New Testament, *bi-* occurs, for instance, with *ner* ‘woman’ where it means ‘queen’, with *Mi* ‘child, son’ only when it is the ‘Son of God’ (capitalization in orthography has a function similar to the uniqueness marker), and with *Uram* ‘voice, word’ only when it is ‘God’s voice’. (13) illustrates its use with a person name as opposed to *a-* with common nouns. Note that Una also can mark person names with preposed third person pronouns not distinguishing gender (*Er Jesus* ‘he Jesus’) and that the two strategies can be combined, as exemplified in (13).

- (13) Una (Nuclear Trans-New Guinea, Mek; New Testament 41009020)
Ba, sun-ci a mi a-si Er Jesus dam bo-ya-nmai
 but they-ERG this child N.UNI-ACC 3SG Jesus near carry-come-PST.3PL
ura, a mi wek-am-we isa a-ryi Er Jesus
 after this child enter-previously-PST.3SG ghost N.UNI-ERG 3SG Jesus
bi-si asing eib-mou ura, a mi a-si tomob-oka
 UNI-ACC eye see-PST.3PL after this child N.UNI-ACC erect-CVB

oublob-mou.

crush-PST.3SG

‘And they brought this child to Jesus. And when the ghost who had entered the child saw Jesus, the spirit immediately convulsed the child...’

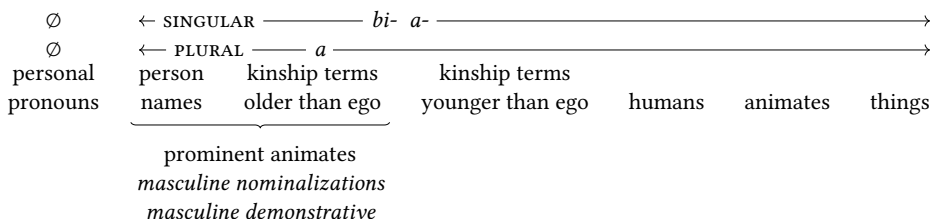


Figure 2: Una *bi-* unique and *a-* non-unique and the animacy hierarchy

The Una gender system can thus largely be characterized by the animacy hierarchy in Figure 2. There are two anomalies (in italics in Figure 2). There is a masculine singular nominalizing suffix *-nyi*, often used in indigenous names, and nominalizations suffixed by *-nyi* are always unique *bi-*.⁴ The same marker *-nyi* can also be added to the demonstrative *a-*, which then together with *bi-* unique can serve as a masculine grammatical anaphor (but is different from the third person pronoun which does not distinguish gender). These anomalies are the germs for a further development of the Una unique marker *bi-* towards a lexical masculine gender *be-* in Nalca (see Wälchli 2018 for the details).

In a wide range of languages from different places in the world, noun markers, whether they distinguish gender or not, are so called PRONOMINAL ARTICLES (Himmelman 2001: 838), which means that noun markers have the same form as personal pronouns (mainly third person pronouns, but occasionally also second and first person) and have developed from personal pronouns (unlike Romance, where both articles and personal pronouns independently originate from demonstratives). For the development of pronominal articles from personal pronouns in Kxoe (Khoe-Kwadi, Khoe), see Heine & Reh (1984: 231–234). Interestingly, many languages with pronominal articles with gender, such as the Khoekhoe language Nama (Khoe-Kwadi, Khoe), Khasi (Khasian, Austroasiatic), Mian and Oksapmin (Ok-Oksapmin, Nuclear Trans-New Guinea), Abau (Sepik), Kayabi and Tenharim (Tupian, Tupi-Guarani), use pronominal articles with proper names (except in vocatives and non-referential use where a person is given a name), in contrast to articles from other origins, which are rarely used with proper names. This

⁴We will return to Mek nominalizations in §7.3 in the discussion of complex controllers.

suggests that person names may play an important role when pronouns extend to articles, and a reasonable hypothesis is that pronominal gender can expand to nouns by travelling down the animacy hierarchy, among other things via person names. Evidence that pronominal articles travel down the animacy hierarchy comes from languages where pronominal articles are less grammaticalized and restricted to human or animate referents. In Oksapmin (Loughnane 2009: 178–184), pronominal articles occur with specific human referents, but usually not with things or animals, where the definite article (of demonstrative origin) is used (14).

- (14) Oksapmin (Nuclear Trans-New Guinea, Ok-Oksapmin; Loughnane 2009: 180)

robin ux=nun bəp ulxe ap jox
 Robyn 3SG.F=OBJ SO 3SG.F.REFL.POSS house DEF/DEM
o=m-de-pti
 leave=PROX.OBJ-make-IPFV.PL.PRS

‘After that, we left Robyn at the house.’

The Oksapmin definite article (a demonstrative) may co-occur with the pronominal article and the order is then noun-definite.article-pronominal.article: *nap jox ux* [younger.sister DEF/DEM 3SG.F] ‘the younger sister’ (Loughnane 2009: 128). In Oksapmin, pronominal articles can be used with animals, when a specific animal is opposed to another one, or for mythical animals with human-like characteristics. Pronominal articles can also occur with forces of nature. There does not seem to be any lexical gender in Oksapmin. Feminine is restricted to female human referents. Oksapmin thus provides support for the hypothesis that, if expressing gender, pronominal articles mark referent-based gender at first, and can later turn into markers of lexical gender, if their extension to inanimate nouns is more advanced.

3.6 The relationship between referent-based and lexical gender

The preceding sections might have evoked the idea that referent-based and lexical gender are strictly opposed to each other, but this is actually not the case. To the extent that lexical concepts denote sets of referents that are homogeneous with respect to the referent-based properties distinguished in gender, there is no mismatch.

The best-known mismatches are so-called HYBRID NOUNS, such as German *Mädchen*, which is neuter as a lexical noun (due to its diminutive suffix *-chen*,

which morphologically assigns neuter gender to the lexeme), but refers to female beings. In cases of conflict, lexical gender is more likely in local than in distant agreement, where semantic agreement (referent-based gender) prevails, which is Corbett's well-known Agreement Hierarchy (15).

- (15) The Agreement Hierarchy (Corbett 1991: 226)
attributive < predicate < relative pronoun < personal pronoun
ein nettes [N] *Mädchen* (N), *das* [N] *ich kenne*. *Sie* [F]...
'a nice girl whom I know. She...'

We assume here that lexical gender is the special case and referent-based gender is the rule. Lexical gender need only be invoked if gender in a language cannot be captured in terms of the animacy or individuation hierarchy. If a language distinguishes marking associated with person names as opposed to common nouns, such as Tagalog, discussed in §3.5, there is no need to invoke lexical gender.

Now, many languages with lexical gender still have choices of gender values that are reminiscent of referent-based gender. Thus, Swedish (Indo-European, Germanic) makes a distinction between the mass noun *öl* 'beer' (neuter gender) and *en öl* 'one.CM beer' (common gender) when it is countable as a glass or a bottle of beer. This distinction is well in-line with the individuation hierarchy, but it is also lexical, since most Swedish nouns denoting liquids do not follow the same pattern.

Plains Cree (Algic, Algonquian, Cree) *mistik* means 'tree' when animate and 'stick' when inanimate (Wolfart 1973: 22; similarly in Cheyenne, see Straus & Brightman 1982: 128). This is again in accordance with the animacy/individuation hierarchy. However, it can hardly be avoided to specify this distinction in a lexical description of Cree and Cheyenne.

If lexical gender develops from referent-based gender, as we assume here, it has to be expected that there are many such cases where the transition from referent-based to lexical gender is tangible. This does not mean that the direction of diachronic change will always be referent-based > lexical. (It is not unlikely that the Swedish, Danish and Norwegian use of counting 'beer' in common – or masculine – gender is an innovation.) However, hybrid nouns re-instantiating the animacy/individuation hierarchy testify to the relevance of referent-based categorization in gender even in languages with predominantly lexical gender.

Audring (2009) shows that pronominal gender systems (where gender is restricted to pronouns) are generally semantically organized; Wälchli 2019 [this volume], in a typological study based on parallel texts, argues that gender in anaphoric use can be addressed in terms of semantic core only, and Bosch (1988: 227)

claims that the descriptive content of gender is activated in contrastive use in implicit or explicit focus. While the development of referent-based to lexical gender can entail loss of transparency (see §6.3), transparency can also be reintroduced, especially in certain anaphoric uses, such as contrastive focus constructions and reference tracking after long stretches of discourse (in line with Corbett's Agreement Hierarchy). Seifart (2018: 24) discusses the case of nouns denoting animals in the Miraña variety of Bora (Boran) that have undergone re-classification to a transparent class. Based on Bosch's (1988) findings, Seifart (2018) argues that contrastive use and other contexts, where the descriptive content of gender is activated, is more frequent with animate than with inanimate nouns. As a consequence, animate nouns are more likely to undergo re-classification to a transparent class. The preference of animate referents for transparent gender is well in line with Dahl's (2000a) findings about the interaction of gender and animacy discussed in §3.2 above.

While there are languages with referent-based gender only (such as Una), there are probably no languages that only have lexical gender and no referent-based gender. Languages and language varieties with referent-based gender only are not restricted to emergence of gender, but also occur where gender is in decline. Modern English is a good example of a language with referent-based gender which developed from an earlier stage with predominantly lexical gender. Notably in cases of intensive language contact, gender systems tend to be reorganized based on animacy, as shown, for instance, by Karatsareas (2014) for varieties of Koineic Greek in Asia Minor (see also Di Garbo & Miestamo 2019 [this volume]).

In many languages, definiteness and referentiality play important roles for whether or not gender is marked. Greenberg (1978) shows how important the definiteness hierarchy is for the evolution of gender systems along the cycle of the definite article (0 demonstrative > I definite article > II non-generic article > III general noun marker). In several languages, gender markers or noun classifiers are missing when nouns are used in predicative function, where they are non-referential (see Fedden 2011: 110–111 for Mian and Grinevald Craig 1977: 330 for Jacaltec [=Popti']).

3.7 Gender and reference tracking

Above we have emphasized the importance of reference for gender, especially from a developmental perspective. One important thing that remains is to show that this does not entail that gender is suitable for reference tracking (see Kibrik 2011: 355 against Heath 1975 and Foley & Van Valin 1984, chap. 7). Much of the

complex ways in which gender deals with reference originates from the fact that it is more important for gender how objects are categorized than what they refer to. Here we will show that this holds both for lexical and referent-based gender. Let us consider lexical gender first.

It is directly understandable why lexical gender is not particularly suitable for reference tracking, since classes of noun lexemes are easily affected by reconceptualization of referents in discourse (also called recategorization). The best known case is probably Cornish's (1987: 256) example from French (see also Croft 2013: 121), where *le potage* 'soup(M) (refined term)' is later referred to by another interlocutor by the anaphor *elle* (F), implicitly associated with *la soupe* 'soup(M) (common term)'.

Similarly, in (16) from Meskwaki (Algic, Eastern Great Lakes Algonquian), where gender is marked on verbal pronominal indexes and free pronouns, a referent is first implicitly construed as 'game (venison)' and later as 'birds'. *Mi:čipe:h-i* [game-INAN.SG] is animate when indefinite, but animate *mi:čipe:h-a* [game-ANIM.SG] when definite "by convention" (Thomason 2003: 380). Here, "by convention" simply means that we are dealing with lexical gender. As in the French example with the soup, the noun is not explicitly mentioned, but only latently present in association with the inanimate form of the indefinite pronoun *ke:ko:h-i* [something-INAN.SG]. The speaker then further specifies the referent as 'birds', a noun that is always animate in Meskwaki, whether indefinite or definite, and therefore the next verb 'he put them' agrees for animate object. This and other examples in Thomason (2003) show that gender does not necessarily remain constant in cases of coreference. However, person, in (16) obviative, is more constant in this respect.

- (16) Meskwaki (Algic, Eastern Great Lakes Algonquian; Thomason 2003: 380)

Ke:ko:h-i

something-INAN.SG

ne:hto:-čini,

whenever.3.ANIM.SG.PROX.killed.it-3.INAN.SG.OBV.ITER

wi:škeno:he:h-ahi nekotah-meko

little.bird-ANIM.PL.OBV somewhere-EMPH

e:h-as-a:či

i:na

kwi:yese:h-a.

3.ANIM.SG.PROX.put-3.ANIM.PL.OBV that.SG.PROX boy-ANIM.SG.PROX

'Whenever he killed anything, birds, that boy put them in a certain place.'

However, keeping markers constant with the same reference does not only fail with lexical gender. This is because a different position on the animacy hierarchy

does not always entail a different referent, as can be seen in (17) from Tagalog, where the proper name *Maria* and the common noun *ina* ‘mother’ have the same referent, but take different topic markers, since the former is a person name and the latter a common noun.

- (17) Tagalog (Austronesian, Central Philippine; New Testament 40013055)
Hindi ba si Maria ang kanya-ng ina...
 not Q PN.TOP Maria CM.TOP POSS.3SG-LNK mother
 ‘Is not Maria his mother?’

In gender systems where person names play an important role, such as Uduk (Koman), Nalca and Owa (Austronesian, Oceanic; Mellow 2013), coreference does not play any major role in agreement. For Uduk, see Killian (2019 [in Volume I]). Note also the gender recategorization in (5) from Mopán Maya in the context of a change of agentivity in the referent.

Following Kibrik (2011: 334–360), Nichols (2019 [in Volume I]) argues that the usefulness of gender in reference tracking is marginal. She argues that gender, unlike person, never refers. If a category is referential, like person, it is the category itself that refers, and not the word that carries that category. Gender can be referent-based (and very often is, as shown in this section), but not referential.

3.8 Conclusion

We can conclude that hierarchical patterning plays an important role for limiting the complexity of gender, but the potential for hierarchical patterning is strongly limited in gender. It is a powerful decomplexifying mechanism only for the top segment of the animacy hierarchy. In this section we have discussed several cases where it can be shown or at least be made plausible that gender originates as referent-based gender in the top segment of the animacy hierarchy. Referent-based gender then tends to leak into inanimate referents due to such factors as agentivity, salience, purview, uniqueness, and possession, which have the potential of linking certain inanimate referents with animacy. These connections cannot easily be arranged on the animacy hierarchy and, as a consequence, hierarchical patterning breaks down. The only alternative, then, to reconstitute order is to organize gender in terms of lexical nouns, and the outcome of this development is lexical gender. Lexical gender cannot easily be organized in hierarchical terms, which means that gender turns into a category that is fully dependent on the part-of-speech nouns. This leads us to the next section, where nominal morphology is discussed.

4 Gender as a special case of the accumulation of nominal morphology

4.1 Introduction

Many gender markers occur within the noun phrase. According to the Canonical Approach, local agreement (i.e., agreement within the noun phrase) is most canonical (Corbett 2006: 21; Audring 2019 [in Volume I]). In this section we will look at gender markers in the wider context of non-lexical markers within the noun phrase, for which we use the term **NOMINAL MORPHOLOGY**.

In the simplest possible noun phrase grammar, the head noun and its modifiers are unmarked. However, the elements of noun phrases tend to accrue markers in languages of most different kinds, and presence of nominal morphology is obviously more complex than its absence. Even if there is a set with only one marker, an inventory of one is still larger than an inventory of zero. Nominal morphology can consist of uniform markers not distinguishing gender, such as the suffix *-pela* in (18) from Tok Pisin (Pacific Creole English; from English *fell-ow*), or gender-number markers, as the plural proclitic *ki=* in Pnar (Austroasiatic, Khasian) in (19) (opposed to *u=* masculine singular, *ka=* feminine singular, and *i=* diminutive/neuter singular), or the feminine gender marker *n(a)* in Bari (Nilotic, Eastern Nilotic) in (20) (opposed to *l(ɔ)* masculine). Accumulation of nominal morphology, including gender marking, does not seem to correlate with high overall morphological complexity (see Nichols 2019 [in Volume I]). As the clitics in Pnar illustrate, nominal morphology need not consist of affixes. Two of the languages used here for illustration, Tok Pisin and Pnar, have low overall morphological complexity.

- (18) Tok Pisin (Pacific Creole English; Verhaar 1995: 417)
Dis-pela kantri Nimrot i bos-im i gat tri-pela bik-pela
 this-NOMIN country Nimrot 3 rule-TR 3 have three-NOMIN big-NOMIN
taun.
 town

‘This country (that) Nimrod ruled over had three big towns.’

- (19) Pnar (Austroasiatic, Khasian; Ring 2015: 339)
ki=ni ki=so ŋut ki=k^hlawat̚ (ki) wa jap jɔŋ u=daloj
 PL=this PL=four CLF.HUM PL=warrior PL NMLZ die GEN M=Daloj
 ‘those four warriors of the Daloi who died’

- (20) Bari (Nilotic, Eastern Nilotic; Spagnolo 1933: 396)
 “Āso *narakwan n-io’ na jɔndya nan ni.*”
 well wife/female F-POSS.1SG REL.F bring.DETR 1SG here
 ‘Well, my wife whom I brought here.’

(18)–(20) illustrate the attributive use of modifiers with their markers. However, nominal morphology can also occur independently without a noun head. For instance, English adjectives in NPs without nominal heads, which take a prop-word *one(s)* (*the big one(s)*; Jespersen 1949: 245–271), and so called free or headless relative clauses (*she who will read this chapter*). In all cases of NPs with modifiers without noun heads, we will speak here of INDEPENDENT use, a term suggested to us by Martin Haspelmath. Instead of attributive markers we will speak of ADNOMINAL markers, since we need a term that also includes markers accompanying the head noun where there is no attribute in the NP. We will therefore speak of independent adjectives (*the big one*) and independent relative clauses (*he who came*) as opposed to adnominal adjectives (*the big house*) and adnominal relative clauses (*the man who came*).

Given the importance of agreement in traditional approaches to gender, the focus of investigation in the typology of gender has been mainly on adnominal markers. However, there is reason to believe that markers on independent elements (in NPs without head noun) are very important for the development of nominal morphology (including gender markers). For the dynamic approach adopted in this chapter it is therefore essential to consider nominal morphology on independent elements to the same extent as markers on NP attributes. The approach adopted in this section is thus more comprehensive than the study of gender usually is in two ways: (i) markers on independent NP-elements (in NPs lacking head nouns) are included as much as markers on NP attributes, and (ii) sets of markers with one member, such as Tok Pisin *-pela* in (18), which do not partition nouns into noun classes and where there is thus no gender agreement, are also included.

In many languages, nominal morphology first develops in independent use and may then eventually expand to attributive use. English relative clauses illustrate this point neatly. When Middle English had lost gender in relative pronouns – Old English *se* M, *seo* F, *þæt* N were replaced by indeclinable *that* in the 13th century – a human/general gender distinction *who/that* was reintroduced in Standard English⁵ from independent relative clauses, probably first with such non-canonical antecedents as personal pronouns as in (21) (Fischer et al. 2000: 91–93).

⁵Herrmann (2005) shows that virtually all British English dialects are less constrained.

- (21) Middle English (Indo-European, Germanic; Wooing Lord 275.18; Fischer et al. 2000)
hwam mai he luue treweliche hwa ne luues his broðer
'whom can he love truly, who(ever) does not love his brother'

In this section we will discuss the following four hypotheses, which are all closely connected:

- (a) Nominal morphology – including gender markers – tends to develop in independent use and therefore there are usually not more markers in attributive than in independent use.
- (b) Many languages have more than one set of markers (see also Dahl 2000b), such as Standard English three genders in pronouns and human/non-human in relative clauses.
- (c) Relative clauses can play an important role in the development of gender.
- (d) The nature of a set of markers has properties from the function where it originates. For instance, distinctions originating from interrogative pronouns are typically human vs. non-human or animate vs. non-animate.

Hypothesis (a) is inspired by Lehmann (1982) and Heine & Reh (1984: 233; based on Kxoe) and is akin to a universal proposed by Edith Moravcsik: “No noun phrase constituent carries more gender, number, and/or case inflection in adnominal use than it does in pronominal use” (Moravcsik 1994; Universals Archive no. 1733; our term for “pronominal” is independent). (d) is inspired by Croft (1994), who argues that different kinds of classifiers and noun classes tend to express different functions. Numeral classifiers tend to express animacy and shape. Noun classifiers and gender tend to express animacy and sex. Possessive classifiers (at least in Oceanic) tend to express edibility.

4.2 A notation system for adnominal and independent marking

In this section we introduce a simple notation system for markers, which is illustrated in Table 1 for English. For each attributive, head, or independent content element in the NP (henceforth in this section simply called *element*), the set of markers used is represented by the number of markers opposed to each other (for the sake of simplicity we count singular values only). Where there are several different sets with the same number of values, these are distinguished with

lower case letters of the alphabet. Thus, “3” in the English independent NP slot stands for *he/she/it*, “2a” in the relative clause slot (Rel) stands for *who/that*,⁶ and “1b” stands for the prop-word *one(s)* in independent adjectives and independent interrogatives (*which one(s)?*). English has two other sets with two values in independent question words (*who/what* as opposed to the adnominal question word *which*, which can also occur independently with the prop-word *which one*), and in independent indefinites (*someone, -body* vs. *something*),⁷ and there are at least four different sets with one value (definite article *the*, prop-word *one*, complementizer [Cmpl] *that*, and genitive 's). We distinguish two kind of possessors. “Gen” is used for noun possessors and “Poss” for pronominal possessors (adnominal *my*, independent *mine*).

Table 1: English nominal morphology. 1a: *the* (indefinite *a/an*), 3: *he/she/it*, 1b: *one(s)*, 2a: (*he/she*) *who/that*, *which*, 1c *that*, 1d 's, *mine*, 2b *who/what*, 2c *-one, -body/-thing*.

	NP	Dem	Num	Adj	Rel	Cmpl	Gen	Poss	Int	Indef
ADNOMINAL	1a	0	0	0	2a	1c	1d	0	0	0
INDEPENDENT	3	0/1b	0	1b	(3+)2a	1c	1d	1d	2b/0/1b	2c

Human/non-human distinctions restricted to relative clauses tend to be disregarded when gender is discussed. Estonian (Uralic, Finnic), which is not usually considered a gender language, has extended the animacy distinction from free relative clauses in Finnic to attributive relative clauses (*kes* ‘who’, *mis* ‘what’), whereas Finnish retains an omni-purpose attributive relative pronoun *joka* and makes the animacy distinction only in free relative clauses (*kuka* ‘who’, *mikä* ‘what’). The approach applied here can be used to get a better grip on nominal morphology falling into classes (sets where the number of markers is larger than one).

The notion of *gender system* might suggest that nominal morphology in a language tends to be uniform in a language with gender or that there are at least not two different sets of markers with a number of items higher than one. English alone shows that this is not the case. However, we do not want to argue that English has more than one gender system or that Estonian has gender.

Even though one might be inclined to believe that interrogative pronouns

⁶Non-restrictive *which* is not a value of its own and is not counted, and adverbial contexts such as *where* are disregarded here for the sake of simplicity.

⁷-*one* in indefinite pronouns is by the way the source for the prop-word in independent adjectives (see Rissanen 1997).

‘who?’ vs. ‘what?’ are very obvious potential sources for animacy-based gender distinctions, interrogative pronouns do not seem to have developed into a full-fledged gender system anywhere as far as we know. This is perhaps because relative pronouns, which can develop from interrogatives and give rise to NP-internal agreement as in English, are largely restricted to European languages (see, e.g., Comrie & Kuteva 2013). There is thus good reason to exclude interrogative-based animacy distinctions in relative clauses from the definition of gender, as well as animacy-based distinctions in indefinite pronouns. However, it is still useful to have a more comprehensive approach to nominal morphology side-by-side with the gender system approach, since interrogative-based relative clauses are, among other things, instructive for how independent markers can interact with adnominal markers, and this may be relevant for gender as well.

In dealing with nominal morphology in general, we need not be concerned with the question to what extent gender builds uniform systems. Many languages, such as English, have more than one set of markers at the same time. It makes sense to have this more general perspective alongside the more focused gender system perspective.

From the point of view of complexity in NP structure, gender agreement is part of a broader phenomenon of marker accumulation. This is why it is important to also consider sets with one marker, as Tok Pisin *-pela* in (18). Given the frequent origin of markers on adnominal elements from markers on independent elements, markers on independent elements cannot be disregarded. For Tok Pisin, there is actually some evidence that *-pela* has originated in independent use. In Australian Kriol, which is related to Tok Pisin, modifiers can be extended with *wan* (‘one’) and *pala*. In Fitzroy Australian Kriol, *dijan* ‘this one’ and *thar-ran* ‘that one’ are obligatory in independent use, whereas *dis* ‘this’ and *det* ‘that’ frequently occur in attributive use (Hudson 1985: 79). In Tok Pisin, the demonstrative *dispela* is not attested without *-pela* in Verhaar (1995). Thus, both sets of single markers in Tok Pisin and sets of two markers in English and Estonian relative clauses suggest that there is a typical developmental pathway from markers in independent elements to markers in adnominal elements, and this suggests that it might be useful to pay more attention to independent elements in studies of gender as well.

In mature gender systems we expect the same kind of markers pervasively entrenched in all adnominal and independent forms – this is what is usually called a gender system with maximum utilization of available distinctive features. The main expected difference is the kind of elements affected. In German, for instance, noun possessors and numerals above two are not affected. There is the ex-

pected animacy distinction in interrogative pronouns, which has also expanded to independent relative clauses (2a), but which has not affected adnominal marking (see Table 2).

Table 2: German nominal morphology

	NP	Dem	Num 2+	Adj	Rel	Cmpl	Gen	Poss	Int	Indef
ADN	3/0	3	0	3	3	1a	1b	3	3	3
IND	3	3	0	3	3/2a	1a	1b/3	3	2a	2b

4.3 Nominal morphology in emergent gender systems

In order to see clear differences between adnominal and independent marking, it may be more promising to look at emergent gender systems, and we will therefore now consider some languages from different families, some of which have figured prominently in the literature on the origin of gender.

Coatzospan Mixtec (Otomanguan; Small 1990: 415) has seven classes for third person pronouns, which occur as stressed free forms and clitics (Table 3), but it is not clear whether the set of genders is a strictly closed class, since some generic nouns also have clitic forms, but, based on the seven rows in Table 3, we label the gender set “7” in Table 4.

Table 3: Coatzospan Mixtec third person pronouns

	Free form	Proclitic form ^a	Enclitic form ^b
Adult	<i>ñaha</i>	<i>ña</i>	<i>ña</i>
Masculine respect	<i>shtaha</i>	<i>shta</i>	<i>shta</i>
Younger masc. man speak.	<i>naha</i>	<i>na</i>	<i>na</i>
Younger masc. woman speak.	<i>chéhnū</i>	<i>chénū</i>	<i>chí</i>
Younger feminine	<i>táhnū</i>	<i>tánū, tá</i>	<i>tún</i>
Animal/spherical object	<i>kiti</i>	<i>kiti</i>	<i>tì</i>
Other inanimate		<i>é</i>	<i>i, Ø</i>

^aProclitic pronouns occur, among other things, in relative clauses.

^bEnclitic forms are used as subjects and objects of verbs.

Table 4: Coatzospan Mixtec nominal morphology. 1a: *é* complementizer; 1b: *iñá* ‘thing’; 2: *sh(o)ó* ‘who, which (anim.)’, *ne(é)* ‘what, which (inan.)’

	NP	Dem	Num	Adj	Rel	Cmpl	Poss	Int
ADN	0	0	0	0	1a/7		0	2
IND	7	0	0	7	7	1a	1b	2
PRED				0				

Example (22) consists of two NPs in apposition with two proclitics. There are heavy constraints in Coatzospan Mixtec on the number of modifiers per NP head. Apposition of NPs is the only option for combining a stressed demonstrative with a relative clause, but the order of NPs in appositional sequences can be freely reversed.

(22) Coatzospan Mixtec (Otomanguenan, Mixtec; Small 1990: 366)

tánū tsikan tánū kishi iku
 F that F COMPL:come yesterday
 ‘that girl who came yesterday’

Attributive relative clauses need not display gender. The marker *é*, which also occurs in complement clauses (probably from *iñá* ‘thing’; de Hollenbach 1995), is always possible and is obligatory, if there is no proclitic pronoun. de Hollenbach (1995), who surveys relative and complement clause formation in Mixtec and Trique languages, argues that the general relativizer and complementizer marker originates in the headless relative function and can be shown to derive from a noun meaning ‘thing’ in many Mixtec languages. The development is not equally advanced in all Mixtec languages. In Ayutla Mixtec, for instance, the complementizer *ñá* (< *ñaha* ‘thing’) is not obligatory in relative clauses.

Given the nominal origin of the markers, the question may arise as to whether relative clauses headed by proclitic pronouns as in (22) could be considered intermediate between headless and strict relative clauses. Such a proposal has been made by Epps (2012) for Hup [=Hupdë] (Nadahup), and Wälchli (2019 [this volume]) argues that many languages have GRAMMATICAL ANAPHORS, which are intermediate between personal pronouns and full noun phrases. In Hup, relative clause heads range from lexical nouns over bound nouns (cannot occur alone in an NP) and classifying nouns to the general dependent suffix *-Vp*, which is why Epps (2012) comes to the conclusion that headedness is best considered to

be a gradient phenomenon. There are many bound and classifying nouns in Hup, which is symbolized by “nnn” in Table 5. Classifiers only very rarely attach to several elements in a row. Thus, example (23) is best considered a sequence of three noun phrases in apposition.

- (23) Hup (Nadahup; Epps 2008: 278)
núp=(g'æt) pihít=g'æt tih=pög=(g'æt)
 this=LEAF banana=LEAF 3SG=big=LEAF
 ‘this big banana leaf’

Table 5: Hup nominal morphology. 1a: *tih* (23); 1b: $-Vp$, 1c: $-n'ih$, 1d: $-n'ih$, 2: ‘who’ vs. ‘what’

	NP	Dem	Num	Adj	Rel	Cmpl	Poss	Int
ADN	0	0	0	0	0		1d	1e
INTERMED	1a	nnn	nnn	(nnn)	nnn		1d	nnn
INDEPENDENT	1a	0	1a	1a	1b/1c	1c	1d	2

According to Epps (2008: 279), Hup can be considered an incipient classifier system. More advanced classifier systems, such as Kilivila (Austronesian, Oceanic), look very much like an expansion of the intermediate area between adnominal and independent use to all functions. In languages with large sets of classifiers, such as Kilivila, it is difficult to apply the notion of independent use. Senft (1986: 81) lists 176 classifiers, of which he could find 92 in actual speech. Only few of them occur frequently and only few of them have translation equivalents with independent forms in languages without classifiers. Numerals for maths, for instance, take either the masculine/people or thing classifier (Senft 1986: 84). However, because demonstratives, numerals and one set of adjectives (24) always take classifiers we have decided to use “nnn” for large set for both adnominal and independent use in Table 6.⁸ Many classifiers are REPEATERS (the noun and the classifier have the same form) or shortened forms of nouns (Senft 1993: 104).

⁸Nauru (Austronesian, Oceanic; Kayser 1993), which has figured prominently in (Dixon 1982: 167), is another Austronesian language with a system similar to Kilivila with many classes, even though not closely related to Kilivila within Oceanic.

- (24) Kilivila (Austronesian, Oceanic; Senft 1985: 379)
M-to-na tau to-paisewa e-tatai ke-veaka kuliga ke-vau.
 DEM-M-DEM man M-work 3SG-cut WOODEN-big rudder WOODEN-new
 ‘This industrious man cuts a big new rudder.’

Table 6: Kilivila nominal morphology

	NP	Dem	Num	Adj	Rel	Cmpl	Poss	Int	Indef
ADN	0	nnn	nnn	0/nnn	nnn		4	1/nnn	nnn
IND	1/2	nnn	nnn	0/nnn	nnn	0	4	2/nnn	nnn

Bora and Miraña – which can be considered two different varieties of the same language – differ from Hup mainly in that class markers are much more frequently used adnominally and in that there is a set of six general class markers (three in the singular: masculine, feminine, and inanimate). Demonstratives, numerals, adjectives and relative clauses in the NP can take either the general class marker, masculine in (25), or the specific class marker, *FLAT*♂*ROUND* in (25). Seifart (2005: 88–100) lists 66 specific class markers and 53 repeaters for Miraña. Many nouns, such as ‘turtle’ in (25), have class markers inherently as part of the lexeme (“nnn” underlined in Table 7, underlined stands for non-inflectional use of the marker). Given the lack of concord, as in (24), noun phrases with several elements can be considered sequences of appositions (see also Passer 2016b).

Table 7: Bora nominal morphology

	NP	Dem	Num	Adj	Rel	Cmpl	Poss	Int
ADN	0/ <u>nnn</u>	3/nnn	3/nnn	3/nnn	3/nnn		0	2a
INTERMED	<u>nnn</u>	nnn	nnn	nnn	nnn		nnn	nnn
INDEPENDENT	3/1a	3	3	3	3	1b	1b	2b/3

- (25) Miraña Bora (Boran; Seifart 2005: 169)
aj:-di/ε:-hi múhuw-hi/múhuw:-be kú:mur-hi
 DIST-M/*FLAT*♂*ROUND* be.big-*FLAT*♂*ROUND*/M turtle-*FLAT*♂*ROUND*
 ‘that big turtle’

Like in Mixtec languages, the marker used in complement clauses (22) is the general class marker for inanimate (“1b” in Table 7) *-nɛ̃/pɛ̃* ‘thing’. It is also used in independent possessors. Note that the whole question in (26) is embedded and *-nɛɛ* ‘INAN/thing’ marks it as a complement clause.

- (26) Bora (Boran; Thiesen & Weber 2012: 364)
ts^ha^{H?} ò kpá:hák^hùr-t^hú [mú-ʔà tsa^S:-nɛ̃ɛ̃
 not I know-NEG who-PL come-INAN
 ‘I do not know who (or what animals) come’, lit. ‘that who(PL) come’

In possession, class marking is limited to the intermediate domain. Possessive pronouns with the inanimate class marker can suffix a specific class marker (Thiesen & Weber 2012: 179).⁹

We have seen above for Mixtec, Hup, and Bora how appositions of noun phrases can contribute to the introduction of markers in adnominal position. Basically there are three possibilities for how attributes and the noun can be connected in the NP: (a) the attribute modifies the noun, which is then its head noun, (b) the noun and attribute are appositions, and (c) there is a HEADEDNESS REVERSAL (a semantic modifier of a phrase is its formal head). Headedness reversal is not equally common for all types of attributes, but is well-known from numerals. In Russian, for instance, numerals higher than four are historically nominalized and the noun counted is in the genitive plural: *p’at’ čas-ov* [five hour-GEN.PL] ‘five hours’ (literally ‘five of hours’). For our purposes it is especially relevant to consider headedness reversals in relative clauses, viz. the so-called head-internal relative clauses or “circum-nominal” relative clauses (Lehmann 1984: 109–121). A gender language with head-internal relative clauses is Mian. Mian has four gender classes (masculine, feminine, neuter 1 and neuter 2; Table 8).

Table 8: Mian nominal morphology

	NP	Dem	Num	Adj	Rel	Cmpl	Poss	Int
ADN	4	4	0/4	0/4	4/0		0	2
IND	4	4	4	4	4	1a	1b+4	2
PRED	0			0			1b	

⁹Aside from third person pronouns with general class markers, there is also an even more general third person pronoun (“1a”) which is used for coreference, glossed as “self” in Thiesen & Weber (2012: 360). This is a further parallel between Bora and Coatzospan Mixtec.

The genders of Mian are distinguished in third person pronouns, which also occur as articles at the end of noun phrases if the noun is used referentially. Adnominal demonstratives replace the article. Attributive adjectives and numerals are not usually followed by gender clitics, but gender clitics may occur with them. Since head-internal relative clauses are noun phrases, they have final enclitic articles, as illustrated in (27). There are also unmarked pronominal relative clauses.

- (27) Mian (Nuclear Trans-New Guinea, Ok-Oksapmin; Fedden 2011: 506)
- nī senso=e Jemeni daak=o walo-Ø-ob=e*
 we.EXCL chainsaw=SG.N1 PN down=N2 buy.PFV-REAL-1PL.SBJ=N1.SG
ayam=o=be
 good=PRED=DECL
- ‘The chainsaw we’ve bought down in Germany is good.’

Mian head-internal relative clauses are sentential nominalizations with the gender-distinguishing article as nominalizer. The same construction can also be used as temporal adverbial clause, but then always has a “neuter 2”-class article (maybe because time nouns are neuter 2).¹⁰

The Ngan'gityemerri variety of Nangikurrunggurr (Southern Daly; Reid 1990) is another language prominently figuring in the literature on the origin of gender. It is like Hup rather than Bora in that class markers on modifiers are optional, and not all nouns belong to a class. Demonstratives and possessive pronouns can stand alone in free use (“0”), but adjectives cannot head noun phrases on their own. Personal pronouns make a masculine/feminine distinction (“2a”), but are mainly used as possessive pronouns (28), since subject and object are indexed on verbs, where gender is not marked (“1”; see Table 9). As in Coatzospan Mixtec, gender markers can occur as free words, as proclitics and as postnominal markers, here suffixes, but only a small number of classes have reduced forms. Eight classes have proclitics: M *wa=*, F *wurr=*, group of people *awa=*, animal/meat *a=*, vegetable *mi=*, dogs *wu=*, tree/things *yerr=*, and *yeli=* bamboo spears. Class suffixation is restricted to the interrogative *tyen-* ‘what kind of’ (“nb”), and the negative particle *minbe-* (“nc”), with which the suffix forms a kind of negative indefinite pronoun. *Tyen-* ‘what kind of’ has a class *-da* ‘country/place’, which does not occur in proclitics. Free interrogative pronouns make a human/non-human distinction (“2b”): *kene* ‘who’, *tyagani* ‘what’.

The same set of markers that are used as proclitics occur as prefixes for deriving nouns (“n”), except bamboo spears, which has the freeform classifier

¹⁰Mian has further a system of six classificatory prefixes on verbs of object manipulation, which is not considered here (see Fedden 2010: 459; Fedden 2011: 185).

Table 9: Ngan'gityemerri Nangikurrunggurr nominal morphology

	NP	Dem	Num	Adj	Rel	Poss	Int	Neg.indef
ADN	0/n/nn	0/n/nn	0/n/nn	0/n/nn	n	0/n/nn	0?/n/nn?	nc
INTERMED	n/nn	n/nn	n/nn	n/nn	n	n/nn	n/nb	nc
INDEPENDENT	1/2a	0	?	n	n	0	2b	nc
PRED				n/0				

yawurr instead. In four classes (animal/meat *a/e-*, vegetable *mi-*, canines *wu-*, tree/things *yerr/yed-*), the clitics have turned into prefixes in some lexicalized forms, and a prefix is also *da/de-* for bodyparts, which is no agreement class. Class marker proclitics can also be prefixed to sentences, so-called GENDERED CLAUSES, such as *a=yenim-walal-pi* [ANIM=3SG.AUX-shake-head] ‘clickbeetle’ (lit. animal-it shakes its head), *a=dudu-meny-tyamu* [ANIM=swollen-3SG:do-cheek] ‘blanket lizard’ (lit. animal-it has swollen cheeks) (Reid 1997: 210). Unlike some other Australian languages, such as Bininj Kun-Wok (Gunwinyguan), where gendered clauses are highly limited in productivity, gendered clauses are fully productive in Ngan'gityemerri, as in (28). Note that the relative clause in (28) actually consists of four clauses with different subjects. The antecedent is possessee of the first, local oblique of the second, subject of the third, and local oblique of the fourth clause. Despite its syntactic complexity, its function is derivational. It serves to express a concept, viz. escalators.

- (28) Ngan'gityemerri Nangikurrunggurr (Southern Daly; Reid 1990: 380; Reid 1997: 205)

yentyi-ngirrki-tye *yerr*=[*watypela* *nem*,
3SG.take-1DU.EXCL.do-PST INAN/TREE=whitefella 3SG.M

wannim-derri-tyerr, *yentyin-yirrimbin*, *wannim-fel* *wun-ambirri*
3PL.go.PRS-back-halt 3SG.take-3SG.go 3PL.go-jump there-ahead

‘He took the two of us onto that thing of whitefella’s, that they stand still on, and it takes them and they jump off ahead there at the top (i.e., escalators).’

All examples of non-lexicalized relative clauses given by Reid have either masculine or inanimate class proclitics. There is a special relative locational marker *ngan-*, but relative clauses do not seem to be used as complement clauses. Ngan'gityemerri has similar sets of markers for different elements, but they have not really grown together into one uniform gender system.

Let us now return to Bari and Pnar, which were exemplified at the beginning of this section. Bari is an Eastern Nilotic language, and the East Nilotic languages have innovated gender agreement with Bari being the language that has the least developed system (Heine & Vossen 1983: 257). Masculine (*lɔ*) and feminine (*na*) gender is distinguished on demonstratives, one type of adjectives, relative clauses, possessive pronouns and noun possessors (except inalienable kin), and the interrogative adjective. Interrogative pronouns, however, have a human/non-human distinction (*ɲa* ‘who’, *nyɔ* ‘what’) (“2b”; see Table 10).

Table 10: Bari nominal morphology

	NP	Dem	Num	Adj	Rel	Cmpl	Poss	Int
ADN	0/2	2	0	0/2	2		2/0	2
IND	1	2	0	0?/2	2	0	2	2b
PRED				0/2				

In Acoli, which is a Western Nilotic language, we can trace the origin of the gender system in a marker set consisting of a single marker. Acoli *lâ-*, PL *lâ-* is a derivational prefix ‘person, individual, one who...’ without distinction of sex, and has the function of introducing gendered clauses which are a kind of headless relative clause: *lâ-ít-é ò-tɔ̀* [NMLZ-ear-POSS.3SG 3SG.PST-die/become.useless] ‘a deaf person’ (Crazzolara 1955: 37). This construction seems to have evolved from NN compounds with *lâ-* as an erstwhile light noun ‘person’ (which is not a noun anymore in Acoli), as in *lâ-bòɲò lě̀èm* [NMLZ-NEG property] ‘one without goods’. The prefix *lâ-* also forms nouns for members of a nation or tribe in Acoli: *Lâ-pàtikò* ‘man of Patiko’ (Crazzolara 1955: 42). Shilluk, another Western Nilotic language, has a similar feminine element *nya*: *nya Lul* ‘a woman/girl of Lul’ (Kohnen 1933: 17). In Acoli, *ny(a)à* means ‘daughter’ and is much more restricted in its use in compounds. According to Heine & Vossen (1983: 263), the “Eastern Nilotic gender markers **lɔ* M, **na* F are likely to go back to lexical items which formed head nouns in genitive constructions”. However, it is important to emphasize that it is attributive possessive construction with non-anchored possessors (possessed expressions without a referential possessor; see Koptjevskaja-Tamm 2005), such as for the expression of membership to a tribe, that we are talking about. Acoli has a different construction for predicative possession with anchored possessors. The non-anchored possessor construction became productive and expanded to possession in general in Bari (but not to inalienable kin), and from gendered clauses, in a similar way as in Ngan'gityemerri, it expanded to relative clauses.

The development probably started with a set with one member **l* M. Feminine **na* joined in later and originally only had the non-anchored possessor function. If, as we assume, Acoli represents the original situation and Bari a secondary development, then this development demonstrates how important it is to include nominal morphology with one member in marker sets if gender is considered from a developmental perspective.

Pnar (Ring 2015) has gender marking third person pronouns and corresponding proclitic noun markers (pronominal articles; *u*= masculine singular, *ka*= feminine singular, and *i*= diminutive/neuter singular, *ki*= plural, “3” because there are three singular classes).¹¹ There are also three numeral classifiers (“3b”) used with numerals above ‘one’: human *ɲut*, non-human *tlli*, and weeks *ta* (see also Sinemäki 2019 [this volume]). Possessors are partly unmarked and partly marked with preposed *jɔŋ* (“1b”), which is obligatory if the possessor occurs without head noun. Relative clauses are nominalized with the preposed marker *wa* (“1a”), which also occurs – together with the gender proclitic – in independent adjectives. Pnar is one of very few languages that lacks an animacy distinction in interrogative pronouns, but the interrogative pronoun *ji* ‘who/what’ and the interrogative adjective *wɔn* or *nu* is combined with the class proclitic. If the gender of the individual or item asked about is known, the appropriate class proclitic is used, otherwise any class proclitic is possible, but *i=ji* is most common then (Ring 2015: 235). However, there is an animacy distinction in an unexpected place, viz. one of two types of adjectives. Type 1 does not take the nominalizer in attributive position (“0”; *ki=sistar tɲmɛn* [PL=nun old] ‘the old nuns’; Ring 2015: 173), but Type 2 requires the nominalizer only when the head noun is human (“2”; see Table 11, example (29)):

(29) Pnar (Austroasiatic, Khasian; Ring 2015: 177): Type 2 adjective

- a. *u=ksaw (wa) hɛʔ*
 M=dog NMLZ be.big
 ‘the big dog’
- b. *u=bru wa hɛʔ*
 M=person NMLZ be.big
 ‘the big man’

Relative clauses always take the preposed nominalizer *wa* (“1a”). Independent relative clauses require a preposed gender marker. Relative clauses with or without a gender proclitic also function as noun complements (19). Interestingly, the

¹¹The personal pronoun in the accusative has different forms.

Table 11: Pnar nominal morphology

	NP	Dem	Num(2+)	Adj	Rel	Cmpl	Poss	Int
ADN	3	3	3b/3+3b	0/2	(3+)1a	3+1a	0/1b/3+1b	3
IND	3	3	3+3b	3+1a	3+1a	0	3+1b	3
PRED				0/3				

gender proclitics also serve for forming various kinds of verbal nouns (*ka*= resultative/F, *i*= inchoative action/DIM, *u*= purposive nominals/M, Ring 2015: 71), as can also be seen in (30): *ka=sərkar* ‘government’ and *u=pŋɛmkam* ‘for using’.

- (30) *ka=sərkar da pŋ=mi? kɔ ki=aŋ [wa ŋ hɔj*
 F=govern REAL CAUS-bring.out 3SG.F.NOM PL=rule NMLZ NEG be.fitting
u=pŋ-ɛmkam plastik]
 NF=CAUS-need plastic
 ‘The government brought rules that it’s not good to use plastic.’

4.4 Nominal morphology in a gender system in decline

Finally, after having considered examples from emerging class systems, let us now look at an instance of a language variety where gender is in decline and which is a counterexample to Moravcsik’s suggested universal that there are never more gender distinctions in attributive than in independent function. We are not making any predictions here about what typically happens in cases of gender loss. However, the example discussed here shows that distinctions in independent use can be lost first, which can result in a system where gender is distinguished only in attributive, but not in independent use.

Whereas Standard Latvian (Indo-European, Baltic) and most Latvian dialects have a fairly canonical gender system with two values, masculine and feminine, Northwestern Latvian dialects [=Tamian], are in various stages of gender loss, which is partly due to Finnic (Livonian and Estonian) substrate (see Wälchli 2017 and Di Garbo & Miestamo 2019 [this volume]). Like Pnar, the Baltic languages Lithuanian and Latvian are exotic in that they lack an animacy distinction in interrogative pronouns (Nau 1999; Lindström 1995). In the dialect of Dundaga, feminine agreement is retained only in attributive function. In all other uses, both independent and predicative, only the masculine form is used. This is illustrated in example (31) with attributive and independent uses of the adjective.

- (31) Dundaga Latvian (Indo-European, Baltic; Dravniece 2008: 87; Wälchli 2017)

Vel' bi [visâ:3 â:d gurĉ] –
 still be.PST all.sorts.NOM.PL.F skin.GEN.SG belt(F).NOM.PL
plattak un šou̯rak, meļ', brūn'
 thicker.NOM.PL.M and thinner.NOM.PL.M black.NOM.PL.M brown.NOM.PL.M
un zēlten'.
 and yellow.NOM.PL.M

‘Moreover, there were all kinds of belts: thicker ones and thinner ones, black ones, brown ones and yellow ones.’

Masculine marking is also generalized in actor nominals, which most typically occur in predicative function: *ou̯d'ēs* [weave.AGN.NOM.SG.M] ‘weaver (of a woman or man)’. See Table 12.

Table 12: Latvian and Dundaga Latvian nominal morphology

Standard Latvian	NP	Dem	Num	Adj	Rel	Cmpl	Poss	Int
ADN	0	2	2	2	1/2	1	2/1	2
IND	2	2	2	2	1/2	1	2/1	1
PRED	2			2				
Dundaga Latvian	NP	Dem	Num	Adj	Rel	Cmpl	Poss	Int
ADN	0	1	2	2	1	1	2/1	2
IND	1	1	1	1	1	1	1	1
PRED	1/2			1				

4.5 Conclusion

Let us now return to the four hypotheses stated at the beginning of this section. Most elements in the few languages surveyed here are in accordance with hypothesis (a) that there are not more adnominal markers than independent ones. However, it is important to point out that the hypothesis does not take into account the intermediate area between independent and adnominal use which is important in classifier languages. Large sets typically develop in the intermediate zone between independent and attributive use as we have seen in the discussion of Hup. This probably holds true also for Bora (synchronically most clearly

in possessives), Kilivila, Mixtec, and Ngan'gityemerri, where the development is more advanced. This means that nominal morphology typically originates in independent or in intermediate function and may eventually expand to attributive use, but not the other way round. It may then occur that a new set in independent use has fewer distinctions than one already entrenched in attributive use, which is the case in German relative clauses, where a human/non-human set originating in interrogative pronouns competes with the three-way masculine/feminine/neuter distinction in relative pronouns. A special case is the Pnar human/non-human distinction emerging in attributive adjectives from the opposition of the presence of the nominalizer *wa* with humans (the same construction as with independent adjectives) versus its absence with non-humans. Finally, Dundaga Latvian shows that in contexts of gender loss the independent function can be innovative in introducing the absence of gender whereas gender is retained in the attributive function.

In this section we have shifted the perspective away from gender systems to sets of markers which need not form systems and entirely different sets may occur in different elements and functions (Hypothesis b). This makes it easier for us to see the many cases where there is arguably more than one set of markers in the same language, which holds for English, Coatzospan Mixtec, Bora, Mian, Ngan'gityemerri, and Pnar. In several cases an animacy distinction originating in interrogative pronouns is involved (English, Coatzospan Mixtec, Bora, Mian, and Ngan'gityemerri). This shows that the question as to what makes a gender system is not a trivial one. Even if marker sets originating from interrogative pronouns are excluded, which is probably reasonable to do, since interrogatives do not seem to be attested as origins for full-fledged gender systems, many languages have more than one marker set (see also Dahl 2000b, who comes to similar conclusions).

We have also argued that relative clauses are important for the emergence of gender (Hypothesis c). Several languages with emergent gender systems have been found to have relative clauses originating from gendered clauses with markers having developed from light nouns. This holds notably for Ngan'gityemerri and Bari. In Mian, a type of relative clauses consists of gender-marked NPs for another reason, namely due to headedness reversal in head-internal relative clauses. This means that relative clauses play an important role in the emergence of gender systems for a number of different reasons.

Can we say in which element and function a gender system originates from its synchronic characteristics (Hypothesis d)? There are at least some trends pointing in that direction. In sets with many markers, some of which express

shape, numerals are usually involved. But maybe more importantly, sets with many markers originate in the intermediate zone between independent and attributive function. In systems where edibility plays a role, possessors are involved. This does not only hold for possessive classifiers in Kilivila, but also for Ngan'gityemerri, where class markers originate from generic nouns. Adjectives and relative clauses have a predilection for marker sets with just one single member. Interrogative pronouns are a frequent source for marker sets with two markers distinguishing animacy. If there is a sex distinction, either anaphoric NP expressions or person name markers or non-anchored possession for the expression of origin (Eastern Nilotic) are usually involved. Independent and intermediate NP sets sometimes make a respect/non-respect distinction as in some Mixtec languages. This suggests that gender does not originate as a full-fledged system, but is shaped by discourse functions in particular local domains.

As soon as independent functions of noun phrase constituents are considered, it is difficult not to have the impression that gender and classifiers lurk behind every corner. To paraphrase Sinnemäki (2019 [this volume]), whose statement is based on a much more respectable sample, languages are more likely to have some noun classification system rather than no noun classification system. Or, in order to say it with Gabelentz (1891), in language, there is always a trade-off between the drive for economy (*Bequemlichkeitstrieb*) and the drive for explicitness (*Deutlichkeitstrieb*). Nominal morphology marking explicitness often develops as a compensation for excessive economy (omission of nominal heads in independent use).

5 Gender assignment

In this section, we discuss systems of gender assignment and consider possible diachronic developments in this domain. §5.1 is concerned with the split between semantic and formal assignment principles, and addresses some shortcomings of this binary typology. §5.2 treats flexibility in gender assignment, and how this phenomenon relates to complexity and the maturity of gender systems.

5.1 Types of gender assignment systems

Corbett (2013) argues that the systems of gender assignment attested in the languages of the world can be subsumed under two main types:

- SEMANTIC ASSIGNMENT, whereby gender assignment is predicted by the meaning of nouns. Semantic assignment systems are further divided into

two subcategories: STRICT SEMANTIC ASSIGNMENT SYSTEMS, where semantic patterns are predictive of the gender assignment of virtually all nouns, and PREDOMINANTLY SEMANTIC ASSIGNMENT SYSTEMS, where for a minority of nouns no clear semantic pattern of gender assignment can be identified. Kannada (Dravidian) is cited as an example of a language with strict semantic assignment: nouns denoting males are masculine, nouns denoting females are feminine, while all remaining nouns are neuter. Bininj Kun-Wok is classified as a language with predominantly semantic gender assignment: gender assignment is largely predictable for most nouns, but certain nouns with similar meanings may be arbitrarily split across two or more genders. For instance, lower animates can be either masculine or feminine, and no clear pattern motivates this distribution. Even though discussed, the two subtypes are not treated independently in the classification and coding system proposed by Corbett (2013).

- SEMANTIC AND FORMAL ASSIGNMENT, whereby for some nouns gender assignment is predicted by their meaning, while for other nouns it is based on formal (phonological and/or morphological) criteria. Semantic and phonological gender assignment is attested in the East Cushitic language Afar. In Afar, nouns denoting males are masculine and nouns denoting females are feminine. For nouns that do not denote sexually distinguishable entities, gender assignment is based on stress patterns: nouns whose unmarked case forms (used, among other things, for direct object; Parker & Hayward 1985: 225) end in a stressed vowel are feminine, and all other nouns are masculine. An example of a language with semantic and morphological assignment is Russian. In Russian, sex is a predictor of gender assignment for nouns denoting males and females. For the rest of the nominal lexicon, gender assignment is predicted by inflectional class. Inflectional classes are in turn defined based on the different patterns of case and number marking that nouns can take.

Corbett's classification of systems of gender assignment is widely accepted in the literature. Yet at least three of the contributions to this work call the classification into question and argue that a bipartite typology does not fully capture the diversity of the gender assignment systems attested among the world's languages. These are Svärd (2019 [in Volume I]) on grammatical gender in New Guinea, Killian (2019 [in Volume I]) on the gender system of Uduk, and the chapter by Dahl (2019 [in Volume I]) on the language ecology of grammatical gender systems. The rationale behind this reappraisal is the same across all three con-

tributions. Both in languages with semantic assignment and in languages with semantic and formal assignment, there are often rather copious portions of the nominal lexicon for which gender assignment cannot be inferred from the meaning of the noun, nor from its formal (phonological and/or morphological) appearance.¹²

While, as mentioned above, Corbett (2013) recognizes that languages with semantic assignment may have residual areas of the nominal lexicon that are not in the scope of the semantic rules which are elsewhere productive, this observation is not operationalized further in his typology. This in turn means that the incidence of arbitrary or opaque mechanisms of gender assignment in the sample used by Corbett (2013) cannot be estimated based on the existing coding. Svärd (2019 [in Volume I]) proposes a revised version of Corbett's typology, where opacity of assignment is one of the criteria at stake, and tests it on a sample of twenty Papuan languages. The analysis shows that introducing a systematic distinction between *Transparent semantic assignment systems*, *Semantic and formal systems*, and *Semantic and opaque systems* provides a more accurate representation of the systems attested in the languages of his sample. While sixteen out of the twenty languages count as displaying semantic assignment systems when using Corbett's classification, these figures drop by half (eight out of twenty) when purely semantic systems are distinguished from systems displaying both semantically predictable and opaque gender assignment.¹³ Similarly, the gender assignment system of Uduk is described by Killian (2019 [in Volume I]) as partly semantic, partly formal, and largely opaque, since for many nouns in the language it is not possible to retrieve any clear-cut connection with semantic and formal assignment criteria. Finally, Dahl (2019 [in Volume I]) suggests that the notion of opacity should be taken into account when studying the diachrony and evolution of gender systems, under the assumption that non-transparent patterns of gender assignment are an indication of highly mature, grammaticalized gender. While the issue of opaque or arbitrary gender assignment is often mentioned in descriptions of individual languages and has occasionally been brought to attention in the general linguistics debate (see for instance Dahl 2000a), this topic has not yet been addressed in large-scale comparative studies of gender systems. In §6.3 we discuss how opaque gender assignment may emerge from, and relate to, semantic gender assignment.

¹²For a recent discussion of semantic transparency and opacity in the diachrony of nominal classification systems, see also Seifart (2018). Opacity is also discussed by Passer (2016b).

¹³Svärd does not exclude the possibility that a language may display a combination of transparent semantic, opaque, and formal assignment, or just opaque assignment. However none of these types is attested in his sample.

Another influential generalization in Corbett's typology of gender assignment systems is that while purely semantic systems are possible and relatively common across languages, purely formal systems of gender assignment are not attested. No matter how important morphological and phonological (or, in principle, opaque or arbitrary) patterns of gender assignment are in a language, there will always be at least a handful of nouns for which gender assignment can be predicted on semantic grounds. This semantic core has been shown to usually target the upper nominal end of the animacy hierarchy, that is nouns denoting humans and (higher) animates, with the cutoff points between these categories varying across languages (Dahl 2000a; see also §3). Killian (2019 [in Volume I]) argues that in the Koman language Uduk, the cutoff point for semantic gender assignment can be higher than 'human'.¹⁴ In Uduk, personal pronouns have inherent gender and are always in class 1; proper names denoting humans (but not place names), on the other hand, are always in class 2. Both personal pronouns and proper names precede human nouns on the animacy hierarchy (1st person > 2nd person > 3rd person > proper names > kinship terms > other humans > animate nouns > inanimate nouns). Below this clearly identifiable cutoff point, semantic predictability in the Uduk gender system is extremely limited.

Di Garbo (2014; 2016) distinguishes between semantic and formal assignment rules and proposes that gender systems with only one type of rule (only semantic) are less complex than systems with both semantic and formal assignment. The relationship between types of assignment rules and the implications for the complexity of gender systems are, however, not discussed in Di Garbo's work. We return to these issues in §6.3 and §6.4, where the relationships between semantic and opaque, and semantic and formal gender assignment are discussed chiefly from a diachronic point of view.

5.2 Flexible gender and the nature of gender assignment

Grammatical gender is traditionally defined as an inherent property of nouns, whereby each noun is lexically associated with only one gender value. Corbett & Fedden (2016: 9) formalize this assumption into the Canonical Gender Principle: in a canonical gender system, each noun has a single gender value. Yet, a moderate to strong degree of flexibility in the patterns of gender assignment that can be productively associated with nominal roots is not uncommon across the languages of the world. In Italian (Indo-European, Romance), many nomi-

¹⁴A similar system is arguably found in Teop (Austronesian; Oceanic). See Dahl (2000b: note 3, 591–592) for details.

nal stems denoting humans and higher animates have a masculine or feminine variant depending on the sex of the denoted entity, cf. *parrucchiere* ‘male hairdresser’ and *parrucchiera* ‘female hairdresser’, where the suffixes *-iere* and *-iera* are productive derivational affixes for the overt marking of gender distinctions. Masculine/feminine doublets for one and the same nominal stem also exist outside the domain of animate nouns. For instance, nouns of trees and the respective fruits often belong to opposite genders, as in *pero* ‘pear (tree)’ and *pera* ‘pear (fruit)’. In a language like Italian, gender assignment is thus exploited as a noun formation strategy, whose interpretation rests either on natural gender distinctions or on other kinds of semantic associations that establish taxonomic relationships or contrasts between entities within a given lexico-semantic field (such as, for instance, names of trees and fruits). For an overview of the relationship between gender assignment and word formation rules, see Contini-Morava & Kilarski (2013). The role of gender assignment in establishing contrasts between semantically interrelated entities is also discussed in §6.3. Gender doublets (as well as triplets and quadruplets) for the same nominal roots are also discussed by (Olsson 2019 [in Volume I], §2.1) for Coastal Marind.

While in all the cases mentioned above the gender contrasts are used to encode different classes of referents within a given lexico-semantic field (male vs. female entities, or types of trees vs. types of fruit), there are languages in which flexibility in gender assignment is used not only for this purpose, but also to express variation in a range of semantic properties associated with one and the same (type of) referent. Consider the examples (32)–(35).

(32) Gitonga (Atlantic-Congo, East Bantu; Carter 2002: 21)

- a. *mu-sankwa*
CL1-boy
‘boy’
- b. *tu-sankwa*
CL12-boy
‘small boy’

(33) Wamey (Atlantic-Congo, North-Central Atlantic; Santos Sachot 1996: 160)

- a. *i-ñí*
CL5-elephant
‘elephant’

- b. *bə-ỹí*
CL18-elephant
'big elephant'
- (34) Tachelhit (Afro-Asiatic, Berber; Penchoen 1973: 12)
- a. *aq-nmuš*
[M]SG-pot
'pot'
- b. *t-aq-nmuš-t*
F-SG-pot-F
'small pot'
- (35) Tachelhit (Afro-Asiatic, Berber; Penchoen 1973: 12)
- a. *t-ay-nžay-t*
F-SG-spoon-F
'spoon'
- b. *ay-nža*
[M]SG-spoon
'big spoon, ladle'

All four examples illustrate instances of switches in gender assignment that are used to encode variation in the size of the noun referent, from default to smaller than default (diminutive) in (32) and (34) and from default to bigger than default (augmentative) (33) and (35). Some crucial differences exist between the gender systems of Gitonga and Wamey as opposed to that of Tachelhit, as well as between their use of flexible gender assignment. Gitonga and Wamey have non-sex-based gender systems with more than five gender distinctions and dedicated diminutive and augmentative genders. Tachelhit has a sex-based gender system with two gender distinctions and no dedicated diminutive and augmentative genders. In this language, switches between the masculine and feminine gender are used to encode size-related types of contrasts. Based on a sample of 84 African languages with gender, Di Garbo (2014) finds that the relationship between type of gender system (in terms of number of distinctions and sex-based vs. non-sex-based assignment) and type of attested size-related gender shifts is rather robust in African languages. Languages with non-sex-based gender and rich inventories of gender distinctions are likely to have dedicated diminutive and augmentative genders, while languages with sex-based systems and a smaller number of

gender distinctions encode the diminutive-augmentative contrast based on the sex-based contrast. In addition to size-related patterns of flexible gender assignment, Di Garbo (2014) finds that gender switches can also be used to modify the countability of nouns; for instance, to form collectives from nouns with regular singular and plural forms or to derive singulative nouns from nouns with collective meanings. Consider the examples:

(36) Eegimaa (Atlantic-Congo, North-Central Atlantic; Sagna 2011: 243)

- a. *e-vval*
CL3-stone
'stone'
- b. *si-vval*
CL4-stone
'stones'
- c. *ba-vval*
CL5-stone
'pile of stones'

(37) Nafusi (Afro-Asiatic, Berber; Beguinot 1942: 32)

- a. *ettefâḥ*
apples(M)
'apples' (collective)
- b. *t-attefâḥ-t*
F-apples-F
'one apple'

In Eegimaa [=Banjal], the regular gender marker for the plural of the noun for 'stone' is *si-* (as exemplified in (36)). However, the noun can be marked by the gender marker *ba-* when the speaker wants to refer to a collection of stones. In Nafusi (37), the masculine, collective noun for 'apples' switches to the feminine gender when speakers want to refer to just one apple. The relationship between gender and the lexicalization of number values is further discussed in §9.

Besides Africa, New Guinea stands out as another documented hotbed of patterns of flexible gender assignment (see also Singer 2018 for an account of flexibility of gender assignment in the Northern Australian language Mawng). These types of systems and their uses have been surveyed by Aikhenvald (2012) and Svård (2019 [in Volume I]), while Dryer (2019 [in Volume I]) digs into the specifics

of the morphosyntax and semantics of the diminutive in Walman, a feature value that in some respects resembles a gender but in other respects does not. Some of the properties of the New Guinean systems, such as the contextual nature of the gender shift and the preferential association between masculine gender and big size and feminine gender and small size, closely match the patterns attested in African languages, and suggest that at least some generalizations about flexible patterns of gender assignment can be made independently of linguistic areas and families. Other properties, such as the existence of dedicated diminutive genders or diminutives reminiscent of gender in languages with sex-based gender, as in the Papuan languages Motuna [Siwai] (South Bugainville) and Walman, seem to be much rarer in African languages, where dedicated diminutive genders are most commonly found in languages with non-sex based gender.

Two questions that can be asked on the nature of flexible gender assignment and that are particularly relevant to the topics discussed in this section are: (i) how does the presence of flexible assignment contribute to the overall complexity of a gender system? and (ii) how can flexible gender assignment be accounted for from a diachronic point of view, that is, from the perspective of the emergence and evolution of gender systems?

The first question has been addressed in work by Di Garbo (2014; 2016), who considers the presence of flexible gender assignment (which she calls manipulation of gender assignment) as a factor that increases the overall complexity of gender systems. This choice is motivated by the fact that in the majority of the languages of her sample, the possibility of manipulating gender assignment as a function of reference construal adds to the lexically specified, inherent, gender of a noun. The noun for ‘boy’ in Gitonga (32) is inherently a gender 1/2 noun and can be assigned to gender 12 when a diminutive construal is intended. The co-presence of inherent and contextual patterns of gender assignment increases the description length of gender assignment rules and thus the overall complexity of a gender system (see our discussion of gender complexity metrics in §2.1). One could argue that not every noun in languages with grammatical gender has an inherently specified gender value. Aikhenvald (2012: 42), for instance, reports that in the Papuan language Manambu (Ndu) only nouns with animate referents have lexical (masculine and feminine) gender, whereas gender assignment with inanimate nouns is entirely referent-based, with the masculine being associated with the encoding of large size and/or long shape and the feminine with small size and/or round shape. Yet, even in a language like Manambu the existence of context-dependent mechanisms of gender assignment combines with the fact that, at least for some nouns (animate nouns) grammatical gender is an inher-

ent, lexically specified feature. Thus, similarly to Gitonga, the co-occurrence of inherent and contextual gender assignment adds to the overall complexity of assignment rules.

Radically contextual gender assignment has been recently documented for Hamar [=Hamer-Banna] (South Omotic) by Petrollino (2016). In Hamar, neither gender nor number marking is obligatory and their occurrence depends on the speakers' choice. Patterns of gender and number agreement are only activated if nouns are overtly marked as masculine, feminine or plural. Nouns in the general form (that is, devoid of overt gender and number marking) do not trigger agreement. In addition, apart from a few kinship terms that have fixed, lexical gender, "any noun in Hamar can be inflected for masculine and feminine grammatical gender, and plural number" (Petrollino 2016: 77). In general, while higher animates display stronger associations between gender marking and the encoding of natural gender distinctions, for lower animates gender marking can also be used to encode variation in size (feminine = augmentative, masculine = diminutive, that is, the opposite of what commonly found in other African languages) and countability (feminine = collective), which becomes systematic with inanimate nouns. Hamar is a rather intriguing instance of a gender system with almost entirely contextual patterns of assignment and non-obligatory gender marking, two properties that would seem to challenge some widely accepted claims about the typology of gender, notably that gender is a lexical property of nouns with obligatory morphosyntactic realization through patterns of agreement. Yet, with respect to the nature of assignment rules, the fact that for a few nouns gender assignment is still fixed suggests that patterns of flexible gender assignment, no matter how radical, would always imply at least some instances of lexically specified gender, and that in sex-based gender systems, lexically specified gender is likely to pattern with humans and higher animates.

Coming to our second question, how to account for flexibility from a diachronic point of view, one could be tempted to assume that highly flexible gender assignment is bound to be more frequent in non-mature gender systems, where a lower degree of grammaticalization prompts stronger referential ties in gender marking, and the gender of a noun is largely determined by the speaker's construal of its referent. While this is a hypothesis that awaits systematic empirical testing, some observations can be made based on already available data.

For instance, we know for a fact that context-based, flexible gender assignment is well attested in highly grammaticalized gender systems such as those of the Bantu and North-Central Atlantic languages, which often have dedicated diminutive and augmentative genders. Since the Atlantic-Congo gender systems

are mature systems, reconstructed in the proto-language and characterized by considerably opaque patterns of assignment, the existence of flexible assignment in these languages would seem to contradict the idea that its presence presupposes young and highly referential gender systems. Interestingly, though, studies of the Bantu gender systems have shown that the dedicated diminutive and augmentative genders (along with the locative genders) are less stable than other gender distinctions and more likely to be replaced by analytic types of evaluative constructions (Creissels 1999; Güldemann 1999; Di Garbo 2014; Verkerk 2014). This could suggest that dedicated diminutive and augmentative genders are less prototypical types of gender distinctions, and therefore likely to disappear or be replaced by other constructions when the system of gender marking undergoes erosion.

Moving on to sex-based gender systems and the use of the masculine/feminine contrast to encode variation in size, countability and/or appreciation/amelioration, one open question is whether the emergence of these patterns of encoding precedes or follows the grammaticalization of a sex-based type of opposition, or whether all these meanings emerge at once, provided that a contrast between two classificatory markers emerge. This issue has been addressed by Mettouchi (2000) for the gender systems of the Berber languages. Mettouchi suggests that the diminutive and singulative meanings of what synchronically is the feminine gender marker *t* developed before the sex-based meaning. According to this proposal, the original function of the marker was purely contrastive. The marker *t* was used to single out an entity with respect to a reference point with which a hierarchical, part-whole type of relationship would be established. The feminine meaning emerged at a later stage and with animate nouns where the pattern of contrast got reinterpreted in terms of natural gender contrast. It remains to be seen whether a diachronic development of this type can be posited for other language families with similar gender systems and uses of flexible assignment.

6 The evolution of noun classes

6.1 Introduction

In accordance with the dynamic approach taken in this chapter, we think that it is crucial to emphasize the diachronic dimension of properties of noun classes. Diachronic developments can be addressed by means of cross-linguistic comparative concepts as much as synchronic systems. Here we will formulate diachronic

cross-linguistic concepts using the formula *From X to Y*. This is all prepared by §3 where we already applied this approach to the developmental path that links referent-based gender to lexical gender, which can be described as:

- (i) From classes of referents to classes of (noun) lexemes (§3)

In this section we will discuss a number of developments in the domain of noun classes. These can be described as follows:

- (ii) From one-to-one assignment to many-to-one assignment (§6.2)
- (iii) From semantic to opaque assignment (§6.3)
- (iv) From semantic to formal assignment (from “covert” to “overt” gender) (§6.4)
- (v) From default genders to gender values with semantic content (§6.5)
- (vi) From classes of single items to classes of larger sets (§6.6)

6.2 From one-to-one assignment to many-to-one assignment

Mature gender systems usually have a limited number of classes. But not all gender systems with two genders are complex. Complex gender systems with a limited number of classes can actually have two different kinds of origins, which seem entirely opposite at first glance. They can develop from many classes or they can develop from two classes. Here we argue that what these two seemingly opposite developments have in common is that there is a shift from ONE-TO-ONE ASSIGNMENT, where every assignment rule applies to another gender value, to MANY-TO-ONE ASSIGNMENT, where the same gender value is the outcome of several assignment rules. This is complexification according to the Principle of One-Meaning–One-Form (§2), as it entails a loss of transparency.

In this section we will consider Dyirbal (Pama-Nyungan) and Khasi, which both have many-to-one assignment. For Dyirbal it has been argued that its four genders have originated from many more classes. Khasi has three genders and they have developed from an entirely transparent pronominal two-gender system with referent-based semantic gender. Despite the entirely opposite range of number of original classes, both developments instantiate the same diachronic comparative concept *from one-to-one assignment to many-to-one assignment*.

Plaster & Polinsky (2007) propose that the four-gender system of Dyirbal (Dixon 1972) has developed from a noun classifier system such as the one described for the rather distantly related language Yidiñ with about twenty classifiers (Dixon 1977), “through the collapse of a larger number of classifiers into a smaller number of genders” (Plaster & Polinsky 2007: 14).¹⁵ Yidiñ and some other Australian languages have classifiers that functionally correspond to the classifiers posited by Plaster & Polinsky (2007) for an earlier, not-attested stage of Dyirbal (see Table 13).

Table 13: Merger of classes to Dyirbal noun classes according to Plaster & Polinsky (2007)

Dyirbal genders and their semantic core	Corresponding classifiers in Yidiñ
I <i>-l</i> : male humans, non-human animates	<i>bama waguḍa</i> ‘male’, <i>mija</i> ‘(edible) non-human animate’
II <i>-n</i> : female humans, birds, stinging things, inanimate nouns related to fresh water or fire	<i>bama buja</i> ‘female’, <i>ḍaruy</i> ‘bird’, <i>buḥi</i> ‘fire, sparks, charcoal, a light, etc.’, <i>bana</i> ‘drinkable liquid’, <i>ḍama</i> ‘stinging animals and plants’
III <i>-m</i> : edible (non-meat) inanimates	<i>mayi</i> ‘edible plant’
IV <i>-Ø</i> : everything else	<i>wira</i> ‘inanimate nouns’, no classifier

According to Plaster & Polinsky (2007), some of the complexities in Dyirbal can be explained by the earlier classifier system that must have been similar to that in Yidiñ. In Yidiñ, a dog “could never be called *mija*” (Dixon 1977: 490). Accordingly, in Dyirbal, ‘dog’ is not in class I, but in class II. For a more limited case of coalescence of homophonous partial repeaters in Boran, see Seifart (2018: 22).

Many-to-one assignment can also develop when masculine and feminine anaphoric pronouns expand and become noun phrase markers for all inanimate nouns, as has happened in Khasi. Rabel-Heymann (1977) proposes 20 semantic sub-classes for feminine, and 14 for masculine nouns. However, the many-to-one assignment does not stop there, since many sub-classes have exceptions. These can even affect compounds. Usually, Khasi compounds have the gender of their head. But although *ka sim* [F bird] and most birds are feminine, some compounds

¹⁵ According to Dixon (1977: 496) it is most likely that both Dyirbal noun classes and Yidiñ classifiers have developed from a smallish set of half-a-dozen or so classifiers.

with *sim* ‘bird’ are masculine, e.g.: *’u sim so’ pho* ‘woodpecker’, put differently, have their own COMPOUND GENDER, where gender of the compound is different from the gender of the head. There are also a number of homophones different only in gender, such as *ka ja* ‘rice’, *’u ja* ‘vegetable’ and *ka dpey* ‘hearth’, *’u dpey* ‘ashes’ (Rabel-Heymann 1977: 271), which demonstrate that gender assignment cannot be phonological here. An example of a sub-class with exceptions is natural forces and landscape features, where 37 feminine items contrast with six masculine exceptions: *’u khnñv* ‘earthquake’, *’u prthat* ‘thunder’, *’u bnaay* ‘moon’, *’u khlvor* ‘star’, *’u l’o* ‘cloud’ and *’u slap* ‘rain’ (Rabel-Heymann 1977: 265).

6.3 From semantic to opaque assignment

According to Svärd (2019 [in Volume I]), opaque gender is characterized by rules that are “not general or have numerous exceptions” (see also §5.1). He finds that 8 of 20 languages of New Guinea in his sample have semantic and opaque gender assignment (as opposed to 8 with transparent semantic and 4 with semantic and formal assignment). Killian (2019 [in Volume I]) describes Uduk as a gender system with largely arbitrary assignment, and Dahl (2019 [in Volume I]) holds that “opaque” or “arbitrary” gender assignment is “a possibility that has been downplayed in recent decades”. Opaque gender assignment is not entirely without rules, but, however the rules are formulated (and there may be conflicting solutions), there are many rules and they make reference to many semantic groups and parts of semantic groups or even to individual nouns. Opaque gender assignment systems do not necessarily lack formal criteria entirely, but non-formal rules (semantic or item-wise) prevail.

Before considering the phenomenon any further, an admonition to caution is in order. There is a risk of adopting extreme positions, on the one hand, by postulating general principles of conceptual underpinning or formal assignment on the basis of a discussion of few examples or, on the other hand, by denying the reality of any assignment rules by emphasizing particular exceptions to trends. Plaster & Polinsky (2007) criticize attempts to explain gender assignment in Dyirbal as semantic by invoking such principles as association in myth or belief, domain of experience, and important property, because they cannot be falsified. “[T]he rules do not apply in any systematic way and, as they are, seem to act more as after-the-fact generalizations than operational principles” (Plaster & Polinsky 2007: 6). They further argue that young children acquiring the language do not have access to the necessary information for motivating gender assignment in such manners. However, their own approach of explaining the gender of some Dyirbal nouns by phonological assignment is not free of after-the-fact general-

izations either. Their five proposed phonological rules only account for one or two examples each, in some cases with as many or more counterexamples, and leave many nouns unexplained.

Here complexity may be useful as a methodological tool. It can be argued, for instance, that if a proposal is not shorter than another one in terms of description length, it cannot be considered adequate. For some languages, such as Dyrirbal, it may be difficult to account for gender assignment in a straightforward manner and often it cannot be excluded that certain generalizations are real (researchers will disagree about them). However, different solutions will share the conclusion that gender assignment in a language such as Dyrirbal is complex in terms of description length.

There are many unrelated languages in all major parts of the world with opaque gender assignment and some of them have developed grammatical gender quite recently, such as Khasi. Hence the question arises as to whether there is anything systematic about these exceptions. We suggest here that these expectations can be summarized as the Principle of Contrast (38), which is an observation rather than an explanation.

(38) Principle of Contrast in opaque gender systems

While nouns in a semantic field often have a preferred gender, some salient nouns in the field tend to contrast with them and take an opposite gender.

Table 14 lists eleven unrelated languages with opaque masculine-feminine gender and how they treat the gender of ‘sun’ and ‘moon’. Difference of gender is in majority, but the null-hypothesis that same gender is equally common cannot be rejected statistically.¹⁶

One difficulty with contrast is that it is unpredictable where exactly there is an opposition in the semantic field (and how semantic fields are delimited). In Paumari (Arawan), a language which “shows a high degree of semantic opacity” (Aikhenvald 2010: 44), masculine ‘sun’ and ‘moon’ are arguably opposed to feminine ‘sky’ (Aikhenvald 2010: 44).

Donohue’s (2004: 334–342) description of Skou (Sko) emphasizes the relevance of gender oppositions and he launches the term “dynamic oppositions”. Oppositions, such as female vs. male, small vs. large, squat-and-round vs. long-and-thin, natural vs. technological, etc., while global, can undergo local reversals, and

¹⁶It is important not to include languages with formal assignment, where the difference in gender need not be accounted for by semantics.

Table 14: Gender of ‘sun’ and ‘moon’ in opaque masculine-feminine gender systems

Different gender	Same gender
‘sun’ M vs. ‘moon’ F: Abau, Skou	both M: Paumari, Rotokas (North Bougainville)
‘sun’ F vs. ‘moon’ M: Bari, Dyirbal, Ket, Khasi, Mian	both F: Manambu, Tunica

“there are different, and contradictory, rationales behind the assignment of feminine and non-feminine gender” (2004: 341). Donohue provides possible explanations for many individual cases of gender choices. For instance, according to Donohue, *tang* ‘canoe’ (feminine), is an extension of land (feminine) and human society (feminine) into the changing, destructive, environment of the sea (masculine). However, what matters for our discussion here is that there are many local gender oppositions which all taken together form a complex pattern, whatever the explanation may be in individual cases.

Dynamic oppositions can especially be observed when two animate genders, typically masculine and feminine, expand below the cutoff point on the animacy hierarchy where gender can no longer be controlled by hierarchical patterning (see §3). Female and male purview (§3.3 iii) will not always be congruent with semantic fields in their entirety, and the metonymic and metaphorical associations at work (which are cross-field rather than intra-field) can be of many different kinds. As a consequence, the expanding masculine and feminine classes have a predilection for sharing semantic fields when taking over them. One reason is that size and shape oppositions, which are frequently observable in opaque masculine-feminine gender systems (see §5; Svård 2019 [in Volume I], and Aikhenvald 2016, chap. 3), easily lend themselves to intra-field oppositions.

Another reason is that cross-domain (metaphorical) associations often have the form of contrasted pairs. According to Capell & Hinch (1970: 49), sun and moon in Australia are always female and male respectively in mythology, even where the local language has no noun classes. However, such oppositions need not always be sensitive to the kind of grammatical gender oppositions at work, which accounts for its unpredictability. In Tunica, where both ‘sun’ and ‘moon’ are feminine, there is still a local opposition, but one of age which is irrelevant for gender: the moon is personified as a granny and the sun as a young woman (Haas 1940: 57).

As discussed by Seifart (2018: 21), increase in opacity can be also due to historical “accidents” when the prototypical referents of a noun change. In Miraña, ‘axe’ takes the classifier *-hi* for ‘flat and round’, since axes were earlier made of stone and round. Similar developments are also attested with pluralia tantum nouns, whose semantically motivated association with plurality may be lost when a noun no longer refers to a multiplicity of entities, but to just one entity, possibly consisting of smaller parts, as in Konso (Afro-Asiatic, East Cushitic) *filaa* ‘comb’ (more on lexical plurality and its relation with gender in §9). As mentioned in §3, opacity can also decrease, for instance, when nouns denoting animals shift to a transparent class in contemporary Miraña Bora (Seifart 2018: 24).

Generally, it may be assumed that animate reference, distant targets and anaphoric use is an attractor for transparent gender, whereas inanimate reference, local targets and non-anaphoric use is an attractor for opaque gender. This is consistent with Audring’s (2009) finding that all pronominal gender systems (where gender is restricted to pronouns) are semantically organized. Not unexpectedly, gender in Uduk (Killian 2019 [in Volume I]), which is highly non-transparent, is local (adjacent) and non-anaphoric. Among the non-transparent gender languages in Svärd’s sample from New Guinea, Ama, with only verbal agreement, is probably the most unexpected case. However, gender agreement in Ama goes with absolutive arguments, which entails frequent use with inanimate referents. Interestingly, in a language with two concurrent gender systems, such as Pauri (Aikhenvald 2010), both systems can be opaque, which adds to their complexity.

We conclude that unpredictable exceptions are the essence of opaque gender assignment. They come into being, among other things, because cross-domain extensions have the form of local, in Donohue’s (2004) terms, “dynamic”, oppositions and by historical accidents (Seifart 2018). Hence in many-to-one assignment in opaque gender, many does not only mean many semantic fields behaving differently, but also local oppositions within semantic fields without any overarching principle. Hence opaque gender assignment is complex, and its complexity may develop rapidly. However, this complexity is not due to the absence of any principles at work, so one could say that its complexity and unpredictability is, to some extent, systematic.

6.4 From semantic to formal gender assignment and from “covert” to “overt” marking of gender

In many languages with gender, the gender of nouns correlates to a large extent with some morphological or phonological characteristics of nouns. The terms

“overt marking” or “formal gender assignment” (Corbett 1991, chap. 3, as well as §5 in this chapter) applied to these phenomena suggest that the morphological or phonological marking is the cause and the assigned gender is the effect. From a diachronic perspective, however, the relationship usually goes the other way round.

There are at least three well-known ways in which overt gender marking can develop:

- (i) As shown by Greenberg (1978) with a focus on African languages, free demonstratives or definite articles indicating gender can fuse with their head noun. For the fused markers, Greenberg further proposes a development from definite article via non-generic article to class affix.
- (ii) When gender markers evolve from repeaters, as in Boran (Grinevald & Seifart 2004: 278–279) or Ngan'gityemmerri (Reid 1997), fused repeaters become gender markers on noun-associated words and can become derivational affixes on nouns. The grammatical and derivational markers will then exhibit a large amount of parallelism given their common source.
- (iii) In languages with many declension classes, most of them will not originally be associated with a gender. In ancient Indo-European languages, only the non-neuter vs. neuter distinction is entrenched in declension classes. The more recent masculine-feminine distinction only correlates with few declension classes (Delbrück 1883: 116–117). However, in many modern Indo-European languages, such as Slavic and Baltic, declension classes strongly correlating with gender have almost completely replaced all other declension classes.

When declension classes disappear, morphological assignment can turn into phonological assignment, which can entail considerable restructuring, as has happened, for instance, in the development from Latin to French (Polinsky & Van Everbroeck 2003) or in Wolof (Atlantic-Congo, North-Central Atlantic; Becher 2001: 46). In Rendille (Afro-Asiatic, East Cushitic), pitch accent on the final mora in the majority of feminine nouns as opposed to masculines with pitch accent on the penultimate mora is due to a lost feminine suffix *-et* (Oomen 1981: 46; Corbett 1991: 102). Malkiel (1957–1958) introduces the term *HYPERCHARACTERIZATION* for the addition of a marker that overtly indicates a category, and a large amount of his examples are about gender in Romance, such as Modern Spanish *cuchara* ‘spoon(F)’ with the addition of *-a* formed from covert Old Spanish *cuchar* ‘spoon(F)’. Newman (1979: 202) argues for Hausa (Afro-Asiatic, West Chadic) that

phonological characterization of feminine nouns has developed by massive application of hypercharacterization. Hypercharacterization is often not transparent in morphology synchronically. Hypercharacterization is thus a pathway from morphological to phonological gender assignment.

Phonological assignment can also originate directly from SANDHI without detour via morphological assignment, as in Nalca (Wälchli 2018). Owa (Austronesian, Oceanic; Mellow 2013: 26) is another example. While the neighboring language Kahua has four classes distinguished by articles (*o* male person names and kinship terms, *ka* female person names and kinship terms, *i* places, and *na* default), as can be deduced from the Kahua Bible translation, Owa has in addition an *e*-class for five kinship terms beginning with *e*-: *ema(-na)* ‘father(-POSS.3SG)’, *ena(-na)* ‘mother(-POSS.3SG)’, *ewa(-na)* ‘older sibling(-POSS.3SG)’, *esi(-na)* ‘younger sibling(-POSS.3SG)’ and *epu(-na)* ‘maternal uncle(-POSS.3SG)’. Borrowed names and words beginning with *e*- are not *e*-class, which shows that phonological assignment is not productive. While the *e*-class is also semantically coherent (kinship terms), kinship terms not beginning with *e*- are masculine, feminine or default. Perhaps *e*- is a relic of an older general person name *a*-class. The closely related language Arosi, also spoken on the Island of Makira, has an article *a* for male and female person names and kinship terms: *a ina-mu* [PERS.NAME mother-POSS.2SG] ‘your mother’ (Capell 1982: 14, 40). Nalca and Owa have in common that gender markers have developed from an extension of person name markers and that the classes of nouns with phonological assignment have very few members.

Güldemann & Fiedler (2019 [in Volume I]) argue that “overt gender” marking on nouns has to be kept strictly distinct from gender, and they term it “deriflection (classes)” as opposed to gender. Some other researchers make the same distinction, but not always using the same terminology. For instance, Evans (2003: 181–221) for Bininj Kun-Wok strictly distinguishes between derivational markers on nouns (he calls this “noun classes”; this is deriflection classes according to Güldemann & Fiedler) and inflectional agreement markers on modifiers (he calls this “gender”, as do Güldemann & Fiedler).

Other grammatical categories, such as number, do not have a restriction that the category has to exhibit syntactic DISPLACEMENT (has to be realized on another word). This is, of course, a consequence of how gender is defined and that other grammatical categories are defined in different ways. For instance, nominal number can be realized both on the head noun of a noun phrase, or, syntactically, on another word in the noun phrase or clause. Here we will argue that the special definition for gender makes sense, even though it differs from how most other grammatical categories are defined, because it is only displacement that turns

gender into a grammatical category. If not displaced, a gender marker is derivational. Gender on nouns changes the meaning of the noun in such a way that it looks more like derivation than inflection. That is, the change in meaning is too large for it to stay within the limits of a lexeme. In this context, Bybee's (1985) notion of RELEVANCE is useful. According to Bybee (1985: 13), a "meaning element is *relevant* to another meaning element *if the semantic content of the first directly affects or modifies the semantic content of the second*" (emphasis in original source). Animacy and sex, which are the semantic core of gender, are directly relevant to nouns. As a consequence, markers expressing meanings related to gender realized on nouns will typically yield derivation rather than inflection, since, as Bybee (1985: 17) puts it, "relevant categories produce derived words that are more distinct in meaning from their bases than the ones produced by less relevant categories, the combinations of relevant notions tend to be lexicalized". For all parts of speech other than nouns, however, gender is not relevant. In Bybee's (1985) scale of relevance categories realized on verbs, gender is the most inflectional and least lexical expression type, which means that gender markers tend to be more distant from the verb stem than markers of other categories.

In this context it is interesting to investigate what happens when function words bearing gender marking, such as articles or demonstratives, fuse with the noun controlling gender. If grammatical gender has to be displaced, it will cease to be grammatical as soon as fusion takes place. This issue is not yet very well investigated typologically, but there are some indications that the prediction holds true at least as a trend. In Bulgarian (Indo-European, Slavic), the definite article is a second position clitic, but behaves differently depending on whether or not it is fused with the controller (Enger & Corbett 2012: 315). If the clitic is realized on another word preceding the noun, such as an adjective, the article expresses the gender of the noun (39a/c). If the clitic is realized on the noun itself, however, its form can be influenced by the declension class of the noun as in (39b/d). According to Dost & Gribanova (2006: 134) the clitic is phonologically clearly an affix when realized on the noun. Inflectional affixes on nouns can lead to complexities in declension classes (see Güldemann & Fiedler 2019 [in Volume I], their term is *deriflection*), but do not constitute additional genders. However, Bulgarian is specific in that the declension class affixes on nouns and the displaced gender markers on attributes have the same function within noun phrases.

(39) Bulgarian (Indo-European, Slavic; Enger & Corbett 2012: 315)

- a. *dobri-jat* *bašt-a*
 good-DEF.SG.M father(M)-SG
 'the good father'

- b. *bašt-a-ta*
 father(M)-SG-DEF.SG.{F}
 ‘the father’
- c. *dobri-jat* *čič-o*
 good-DEF.SG.M uncle(M)-SG
 ‘the good uncle’
- d. *čič-o-to*
 uncle(M)-SG-DEF.SG.{N}
 ‘the uncle’

For similar phenomena in Norwegian dialects see Enger & Corbett (2012), where the situation, however, is complicated by the fact that feminine gender is in decline.

Fusion of an article with its head noun is a syntagmatic process. There is a paradigmatic parallel to this if inflectional markers on gender targets are extended to inflectional marking on nouns, which has happened in Latvian. Indo-European has different inflectional suffixes on nouns and pronouns. In Baltic, the pronominal suffixes are extended to adjectives and in Latvian even to nouns in some case-number forms, especially in the dative singular. This entails that the dative singular nominal suffix in Latvian correlates 100% with gender (Nau 2011). Not only do masculine nouns of the *ā*-declension (usually feminine) have masculine agreement, they also take the masculine ending of the dative singular *-am* rather than the feminine ending *-ai*: *puik-am* [boy(M)-DAT.SG] ‘to the boy’ vs. *mās-ai* [sister(F)-DAT.SG] ‘to the sister’. As in Norwegian, there can be further complications when gender in such a system is in decline (Wälchli 2017).

Since derivational phenomena connected to gender are beyond the scope of Hockett’s classical definition of gender, they are often completely disregarded. An exception is Contini-Morava & Kilarski (2013), who argue, among other things, that “[n]oun class markers and classifiers can be used to expand the referential power of the lexicon either by creating new lexical items or by presenting referents from different perspectives” (Contini-Morava & Kilarski 2013: 263). It is unclear to us whether this means that languages without noun classes and classifiers have less referential power in their lexicon. However, we would like to emphasize here that it also might be argued that gender, especially in languages without formal assignment, can mean that the lexicon can have fewer elements. As mentioned in §6.3 above, Khasi sometimes has a remarkable lexical underdifferentiation just because gender disambiguates as in *ka brrw* [F human.being] ‘woman’ vs. *’u brrw* [M human.being] ‘man’ (Rabel-Heymann 1977: 270). How-

ever, as pointed out by Dahl (2019 [in Volume I]) it is not always quite easy to distinguish between formal and non-formal gender assignment. From a discourse perspective there is nothing particularly covert about gender in Khasi, since articles distinguishing gender are very frequent. Their status as independent word is not entirely clear either, and at least in the closely related language Pnar, the articles are clearly clitics (Ring 2015).

Sometimes decisions whether a language has formal or non-formal gender assignment are quite arbitrary. Algonquian languages are usually considered to have semantic gender assignment (Corbett 1991: 24), which is, however, quite opaque despite its motivation by power, see §3.3 (v). However, the plural marker on the noun clearly distinguishes between animate and inanimate, which is a morphological distinction. In Meskwaki, even the singular is clearly distinguished in noun inflection (see Table 15).

Table 15: Meskwaki noun inflection (Thomason 2003: 10)

	Animate	Inanimate
singular proximate	<i>-a</i>	<i>-i</i>
singular obviative	<i>-ani</i>	<i>-i</i>
plural proximate	<i>-aki</i>	<i>-ani</i>
plural obviative	<i>-ahi</i>	<i>-ani</i>

Gender in Algonquian can thus be said to be both semantically and morphologically assigned, as there are separate morphological paradigms for animate and inanimate nouns, but it is usually assumed that semantic assignment in Algonquian is so pervasive that morphology is secondary.

6.5 The development of non-noun controllers and neutral genders

It is often argued that noun class systems differ from classifiers in that all nouns must have a gender. As we will see in §8.3, this property of gender systems is closely connected with degree of formalization, which goes hand-in-hand with cumulation of gender with number and/or case. From a developmental perspective, this means that obligatorification of gender is not necessarily a gradual internal development within the category of gender, but is connected to the fact that the development of gender tends to be parasitic on other category types, notably number and case, which are already highly grammaticalized by the time gender starts emerging. As a consequence, there are hardly any attested develop-

ments from non-obligatory to obligatory gender. However, this does not mean that all gender systems where all nouns have a gender value are equally mature. Some languages have large default classes, which may be an indication of a non-mature gender system. Such a language is Nalca, where most nouns have default noun gender (see Wälchli 2018).

If all nouns are gender controllers, this usually implies that there are at least some gender targets where there is a forced choice of gender values. This even holds if there are no noun controllers, which means that the gender system must account for non-nominal controllers. This is the major topic of this section.

Many languages have some kind of NEUTRAL agreement form used for agreement with “non-prototypical” controllers, such as infinitive phrases, clauses, interjections and other quoted phrases, where the term “non-prototypical” does not say anything else than that the controller is not a noun. A potential solution is to say that non-noun genders are default genders. However, this is problematic, as argued by Corbett (1991: 214) in languages such as Spanish or Lithuanian, where there is a unique neutral agreement form dedicated to agreement with non-prototypical controllers (*ello* ‘it, that’, *lo curioso de esta situación* ‘the curious thing about that situation; Corbett 1991: 214–215). Both in Spanish and Lithuanian, the unique neutral agreement form is a relic of the neuter gender, that remains after all nouns triggering neuter in Latin and Baltic switched to another gender. So it is arguably at least diachronically a default.

“Default” is usually thought of as last resort, or, as Corbett & Fraser (1999: 71) put it, “the default is the last thing you get to do”. However, neutral agreement can be expansive, which is not easily compatible with default as last resort. This is the case in the so-called “pancake” sentences in Scandinavian languages (see Faarlund 1977 and Enger 2004 for Norwegian). In Swedish and Norwegian, predicative adjectives usually agree with the subject in number and, in the singular, in gender. In (40) there is no such agreement. But (40) cannot be an instance of last resort default, since there is a subject to agree with. If default is conceived of as last resort, then it is strange that such a default can be extended.

(40) Swedish (Indo-European, Germanic)

Pannkakor är gott.
pancake(CM).PL be:PRS good.N.SG

‘Pancakes are good.’

Sentences such as (40) have the connotation of an event, in this case ‘to eat pancakes is good’. In fact, Faarlund (1977) argues that we have to deal with reduced subject clauses, an analysis which is dismissed by Enger (2004), who also shows

that any last-resort default analysis runs into serious troubles. Corbett (2006: 150) speaks of extension in use of the default, but last resort defaults cannot be extended, if there is some other choice. Enger (2004) shows that Norwegian and Swedish pancake sentences are largely semantically conditioned (low individuation) and are subject to Corbett's Agreement Hierarchy. Following Widmark (1966), he also points out that the syntax of the subject NP plays a certain role. In (41) the common gender form (M/F) of the predicative adjective is possible if the subject has an adjective attribute, but is ungrammatical if the subject is a bare noun.

- (41) Norwegian (Indo-European, Germanic; Enger 2004: 24)
Russisk vodka er sunn / sunt.
 Russian.M.SG vodka(M).SG be.PRS healthy.M/F.SG / healthy.N.SG
 'Russian vodka is healthy.'

If neutral gender and in particular its extensions in Scandinavian languages are semantic, this means that non-noun controllers can be as meaningful in classification as noun classes and should not necessarily be considered to be assigned by default.

In a diachronic perspective, this means that if noun classes have become obligatory for nouns in the sense that every noun must trigger a gender under certain circumstances, there will usually also be obligatory gender agreement with some non-noun controllers. A neutral gender may originate as a default, but the existence of unique neutral genders, as in Spanish, and of expansive neutral gender, as in Scandinavian, shows that neutral genders can become phenomena of their own. Put differently, default genders can acquire semantic content.

6.6 From classes of single items to classes of larger sets

Many classifiers and genders have originally a very restricted range of application which can be gradually extended. According to Erbaugh (1986: 428), Chinese classifiers start out as specific for single items both in diachrony and in child acquisition, and several of the commonly used Mandarin classifiers, such as *běn* for 'books' and *dūo* for 'flowers' (also 'clouds'), still tend to be restricted to single or few concepts in spoken Mandarin of adults. The Chinese general classifiers Mandarin *ge* and in earlier periods *méi* have developed from words for 'bamboo' and 'trunk of bamboo tree' and can be shown to have gradually extended their range of application (Erbaugh 1986: 429).

The Ngan'gityemerri 'thing'-class *yerr-* has developed from *yawurr/yewirr* 'tree', "a natural extension of the allocation to this class of such traditional artefacts as woomeras, spears, shields, coolamons etc, which are all made from the timber of trees" (Reid 1990: 309). Similar developments are found in other Australian languages. According to Allan (1977: 300), a classifier for trees and objects is perhaps the most common inanimate classifier.

The Ateker [=Teso-Turkana] group of East Nilotic has developed a third gender (in addition to the East Nilotic two gender masculine-feminine opposition, as in Bari; see §3) from an anaphoric noun **(né)ni* 'that place (just referred to)' (Dimmendaal 1983: 219; Heine & Reh 1984: 228). In Turkana, an Ateker language, "[t]he new gender has lost virtually all traces of its locative origin" (Heine & Reh 1984: 229) and has mostly a diminutive function. In another East Nilotic language, Maasai [Masai], only one noun *wwéji* 'place' belongs to the third gender, which indicates place (Tucker & Mpaayei 1955: 15; Payne 1998: 160).

Not only individual nouns, but also, for instance, person names, nominalizations and diminutives (or particular types of them) can serve as starting points for the development of larger sets of classes. Similar to extensions from the use with particular nouns are extensions from other very specific functions. In Iroquoian languages, a feminine(-indefinite) gender in Five Nations Iroquois has developed from a generic and indefinite human index on verbs (still represented in Cherokee, Southern Iroquoian; Mithun 2014: 141). This development is an extension of referent-based gender rather than of lexical gender. In Mohawk (Iroquoian, Northern Iroquoian), the feminine-indefinite gender is mostly used for older women and for expressing respect, but individual speakers use it for specific sets of referents. "Often mothers are initially shocked to realize that they use different gender prefixes for different daughters" (Mithun 2014: 138).

Many languages have some sort of respect or honorific distinction especially for women. Respect can originally be associated with specific nouns, as in the case of Lak (Nakh-Daghestanian). In Lak female nouns originally were gender II with the exception of *duš* 'girl, daughter' gender III. *Duš* 'girl, daughter' served as a so-called *Trojan horse* (Corbett 1991: 100) to transfer all female nouns to gender III except older family members, because this gender is associated with politeness. In two different Polish dialect areas transitional to Czech and Slovak (Indo-European, Slavic), specific morphological formations for diminutives and patronyms for unmarried women, which happen to be neuter and masculine respectively, are the origins for neuter and masculine gender use for unmarried women (Corbett 1991: 101). As in Mohawk, it may play a role that the communities are small and everybody knows each other for keeping track of the sets of referents.

In some varieties of Swiss German, nicknames for women are either neuter or feminine depending on their morphological pattern. Accordingly, the anaphoric pronouns referring to specific women whom one knows well are either neuter or feminine, and the neuter pronoun *ääs* ‘it’ can be used contrastively. Bosch (1988: 218) argues for Standard German that the neuter pronoun *es* ‘it’ never can be accented, since the marked (emphatic) pronoun reflects a classification of referents. If Bosch’s descriptive-content hypothesis is true, this means that the set of referents is part of the descriptive content of Bernese German neuter and feminine genders. In (42), there are two different women or girls with the name *Susanne* with different diminutive-based nicknames, one of which is feminine (*Suslə*) and one of which is neuter (*Susi*). The referents can be tracked by using the corresponding emphatic forms of personal pronouns in contrastive focus.

- (42) Bernese Swiss German (Indo-European, Germanic; constructed example)
D Suslə v ds Susi sötə beedi choo.
 DEF.SG.F Susle(F) and DEF.SG.N Susi(N) should.COND.3PL both.PL come.INF
SII isch scho daa, abər EES no nid.
 SHE be.PRS.3SG already here, but IT not yet.
 ‘Susle(DIM.F) and Susi(DIM.N) are both supposed to come. SHE is here already, but “IT” is not yet here.’

The Mohawk, Polish and Swiss German examples show that extensions to general sets is not only relevant for lexical, but also for referent-based gender.

Since every noun class has its own history, it will usually be the case that different classes in the same languages are at entirely different levels of generalization. Mandarin, for instance, has a special classifier for ‘books’, *běn*, as opposed to the general classifier *ge*. The Ngan’gityemerri *kurum/kurim* ‘canegrass spear’ class is opposed to the much more general ‘thing’ class *yerr-*. It is hence astonishing that there are languages with a roughly equal distribution of nouns across genders. This is the more likely the lower the number of genders is, since the general tendency for Zipfian distributions is hard to do away with in larger sets of items. According to Zipf’s Law (1935), the frequency of a form is inversely proportional to its frequency rank. This entails that different forms will greatly differ in frequency.

Noun classes with very limited scope are not restricted to early stages of development. It is common that a gender disappears by steadily losing its members until nearly no members are left. Standard Swedish has largely lost the opposition between masculine and feminine in nouns, but retains the distinction for anaphoric gender in personal pronouns, and, to a certain extent, masculine in weak

adjectives (see Dahl 2000b, as well as Di Garbo & Miestamo 2019 [this volume]). Traces of the earlier ability of nouns to trigger feminine are the anaphoric use of the personal pronoun *hon* ‘she’ with the antecedent *människa* ‘human being’, and with *klocka* ‘clock’ (Teleman et al. 1999: 61). However, as will be discussed in §7.7, such a development has typically a component of idiomatization, which is characteristic for relics.

7 The complexity of agreement

7.1 Toward a definition of agreement

In this section we will define agreement as an asymmetric specific relation involving displaced information between a syntactically potentially complex controller and a syntactically potentially complex target. The rest of §7.1 has the aim of motivating this definition. §7.2 deals with specific relationships in agreement, §7.3 with complex controllers and §7.4 with complex targets. In §7.5 we argue that features are a mature form of displaced information. Viewed in the dynamic perspective adopted in this chapter, agreement can gradually develop when nouns decategorialize to gender targets (§7.6). But targets can also have properties of controllers and agreement can be idiomatic, which contributes to the fuzzy character of agreement (§7.7).

Throughout §7 we will use the term *complex* in three rather different, but still connected, senses. As elsewhere in the chapter, we conceive of complexity as absolute descriptive complexity (see §2). Thus, *complex* means ‘non-trivial in structure, so that an exhaustive description cannot be short’. The sheer length of this section will suggest that gender agreement is complex. However, this has much to do with *complex* in the sense of ‘consisting of different, but related phenomena’. The phenomena commonly subsumed under agreement in the typological literature are of many different, but yet related kinds. We will argue here that this can be accounted for by the dynamic approach. Different kinds of agreement or agreement-like phenomena have different kinds of origin and represent different stages of maturity of agreement. Throughout the section, we will emphasize the importance of identifying some kind of unequivocal specific relationship between the elements involved in agreement. As we will show, this requirement can only be maintained if the units linked by agreement can be viewed as elements that can consist of several words (complex targets and complex controllers). *Complex* therefore also has the sense ‘consisting of several elements’. Agreement is also complex in the sense of number of items involved,

which, however, limits its complexity in the sense of transparency. We claim that agreement relations can always be seen as one-to-one relations. This comes at the cost of rejecting the idea that agreement is simply a direct relationship between two words. In gender, the word(s) or context(s) triggering the choice of a grammatical value of a marker do not originate in the word on which the category is marked, which is what we call displacement of information, the very essence of agreement. If information is displaced, this happens in the form of a chain (from one chain link to another) and this chain can have more than two links. As we will show, it is not uncommon that there can be many links in the information transfer chain.

Agreement is one of the most traditional notions in linguistics. However, modern linguists very much disagree about its nature and whether it is a useful concept. While Corbett (2006) understands agreement in terms of the highly abstract notion of morphosyntactic feature triggered by a controller and expressed on a target (a conception compatible with many formal approaches), Croft (2013) questions the usefulness of the concept of agreement and Haspelmath (2018) holds that agreement is poorly defined. Haspelmath suggests to replace the notion of (potential) agreement target in the definition of gender by “noun-associated form”, which he defines as “an adnominal modifier (article, demonstrative, adjective, or numeral), or a verbal argument index (subject or object index), or an anaphoric pronoun”. One of Haspelmath’s arguments is that noun-associated forms are often used in situations when the corresponding noun is not overtly present. This argument has been well-known at least since Barlow’s (1999: 190) discourse-oriented approach to agreement. Barlow conceives of agreement as a process of feature unification, a view shared by Corbett (2006) and approaches within Lexical-Functional Grammar (e.g., Kuhn & Sadler 2007). This implies that gender marked forms need not actually be controlled and can occur on their own, such as the isolated utterance in Spanish *Bella!* [beautiful.F.SG] ‘(You are/she is) beautiful!’ While the unification approach accounts for syntactically non-controlled targets, it has the disadvantage that it models agreement as a symmetric relation.

In our view the concept of feature shared by a controller and a target captures the idea that agreement is a form of displaced information. The form where gender is realized is not where gender originates. The information necessary to determine the gender value comes from another part of the utterance, be it from a word or sequence of words elsewhere in the sentence or discourse or be it from the context.

We think that our definition of agreement has the potential of building a bridge between the extreme positions of either viewing agreement in highly abstract

terms or doing away with the notion of agreement altogether. In our view, morphosyntactic features are highly mature forms of displaced information. This understanding of morphosyntactic features allows us to put agreement into a dynamic perspective, where abstract features can be seen as emergent.

Gender marking is the forced choice from a set of redundant marking options on an element that is typically not a noun (the target). The choice is externally determined by an overt, typically nominal, element (the controller) or by the context. In the latter case we will speak of a latent controller, as in Tasmowski-De Ryck & Verluyten's (1982: 328) famous French example (43), where there is no syntactic antecedent for the anaphoric pronoun.

- (43) French (Indo-European, Romance; Tasmowski-De Ryck & Verluyten 1982: 328)

(John is trying to stuff a large table (*la table*, feminine) in the trunk of his car; Mary says:)

*Tu n'arriveras jamais à la [F]/*le [M] faire entrer dans la voiture.*

'You will never manage to get it into the car.'

In (43), the latent controller is lexical. The displaced feature originates in a specific lexical item in a specific word form (singular) which must be activated in the speaker's mind when she utters the target. The controller is not overtly present in syntax, but there is still displaced information from a latent lexical item to a target and there is an asymmetric relation between the latent controller and the target.

A consequence of externally forced choice (the use of *la* or *le* in (43) is not random) is that the target bears some typically nominal information, which can be some referent-based property, such as animacy, sex, size or shape, or some lexical feature of a noun, as in (43). In gender agreement, this is always displaced information. Displaced information is hence a defining criterion of gender agreement. However, there may be intermediate cases where it is unclear whether the choice of a marker is externally determined (by syntax or discourse), especially if the target is a noun or a noun phrase. As we will see, this is highly important from a diachronic perspective since gender agreement, at least in some cases, can be shown to evolve from the decategorialization of nouns (see §7.6). Agreement, as we conceive of it, is partly fuzzy. On the one hand, there are clear cases of agreement; on the other hand, there are phenomena which only have some properties of agreement, and the latter are important for understanding how agreement evolves diachronically.

Much of the disagreement about the notion of agreement comes from attempts to define agreement uniformly syntactically (e.g. Hengeveld 2012 and Passer

2016b)¹⁷ or uniformly in terms of discourse (e.g., Barlow 1999). This has to do with the fuzzy nature of coreferentiality. Many agreement phenomena can be subsumed under Moravcsik's (1978: 363) Coreferentiality Principle, according to which all agreement targets include reference to the controller nominal. The Coreferentiality Principle is too narrow, as shown by Corbett (2006), since targets sometimes agree with non-coreferential controllers, for instance, if the gender of the possessee is marked on the possessor (see §7.6). However, coreferentiality is neither strictly opposed to lack of reference nor to completely independent reference, and, as a consequence, lack of reference is not strictly opposed to completely independent reference. Attributive adjectives in an NP are arguably coreferential with the head noun of the NP, but since it is the NP as a whole that refers, it is unclear whether reference need be invoked at all in this context. Croft (2013) argues that there is no principled difference between independent reference and dependent reference, which leads him to an independent reference analysis of "agreement" as indexation.

The information displaced through agreement marking can consist of features, such as gender and number, with very limited sets of possible values. Features are a mature form of displaced information, but displaced information is not restricted to features: it can be just a condition (a context where a certain choice of marker is made, whereby that context is information relevant for the choice). Corbett (2006, chap. 6) keeps features and conditions apart. We think that the two are just two different forms of displaced information in agreement and features are often mature conditions (which also explains why features often cumulate, see §9).

Agreement is syntactic to the extent that it involves words or groups of words as targets and controllers, but the relation between controller and target can be semantic and it can be inter-sentential. Our definition differs from Corbett's (2006: 4) in seeing the domain as part of the syntactic target. For instance, a target is not an adjective *qua* part of speech, but an adjective in a specific syntactic environment, such as attributive or predicative.

The visualization in Figure 3 captures some core ideas of our definition of agreement. Controller and target can consist of several words (small boxes within a larger box). They are linked by a specific relation, which can be syntactic or se-

¹⁷Hengeveld (2012), van Rijn (2016) and Passer (2016b) only distinguish between independent reference and agreement as the result of a syntactic feature copying-mechanism not involving reference. In their view, attributive adjectives exhibiting gender are an instance of agreement exactly because they do not refer. This entails a narrow definition of agreement, especially manifest in Passer (2016b: 86), who argues that concordial class systems must have a range of language specific modifiers, i.e. agreement within the NP.

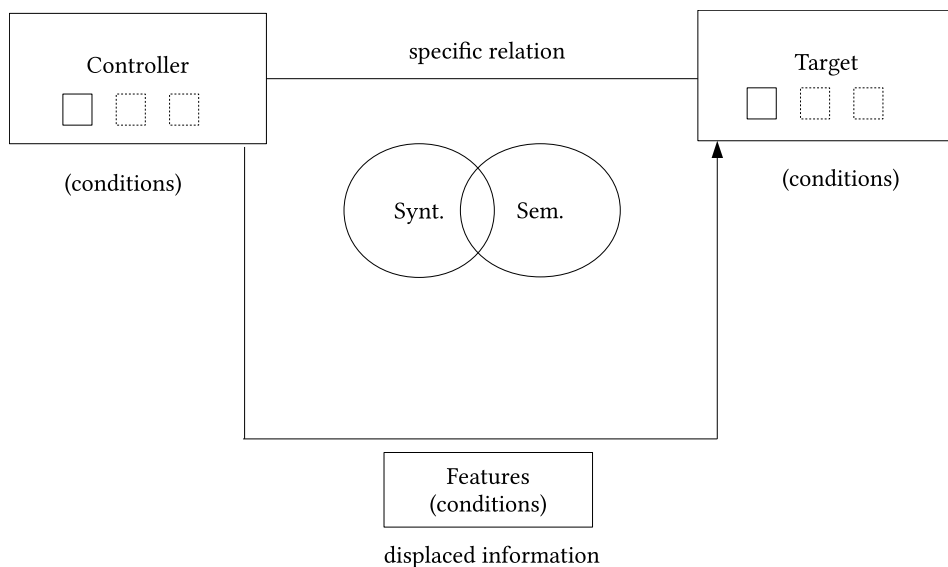


Figure 3: A model of agreement

mantic, and the information expressed by agreement and realized on the target is displaced information, in the sense that it does not originate in the target. It can consist of features or a feature or a condition on a feature.

Unlike Corbett, we think that the notion of relation is an indispensable defining term of agreement. Our definition is consistent in this respect with Lehmann, who requests “a grammatical or semantic syntagmatic relation” between controller and target (Lehmann 1982: 203). However, we think it need not be syntagmatic, since the controller can be contextual. Coreferentiality is an example of a specific relationship that can hold between a controller and a target. Our claim is that there is always some specific relationship between target and controller in agreement. The relationship must be specific, because it must be unequivocal. This does not exclude occasional instances of ambiguity, but agreement is basically a one-to-one relationship or association, not a one-to-many or many-to-many relationship between controller and target. This entails that targets are sometimes complex in the sense that they consist of *FORMAL GROUPS*, a term that we take from Croft’s (2001: 190) critique of the notion of constituent. A formal group can be a phrase or constituent, but it can also be another kind of syntactic grouping (which need not be a continuous string of words). Agreement is a means of indicating syntactic grouping (groups of controller and target or tar-

get groups or controller groups) parallel to, but not necessarily congruent with, constituency.

Sometimes a word displaying an agreement marker does not entertain any specific or exclusive relation with the controller if considered in isolation, but only when considered in terms of a formal group it is part of. This is evidence for syntactically complex targets. Consider (44) from Italian with object agreement in the participle of *dovere* ‘must’ in the verbal formal group *ho dovuti chiudere* ‘had to close’ preceded by the object clitic which triggers the agreement. The object is semantically an object of the verb *chiudere* ‘to close’, but the whole sequence consisting of three verbs is a unit when it comes to argument structure. In terms of Rizzi (1982), the modal auxiliary and the lexical verb form a verbal complex. As further elaborated in §7.4, the verbal complex in (44) is an instance of a complex target or target group. The clitic, which is coreferential with ‘pralines and biscuits’ in (44), is an object of the whole verbal complex, not just of ‘must’, and as an object of the whole verbal expression it triggers object agreement, which happens to be realized on the modal verb, because the participle is the only form in the verbal complex where object agreement can be realized. Furthermore, [*q*ueste *praline e questi biscotti* ‘these pralines and biscuits’ is an instance of a complex controller with gender resolution (in Italian feminine and masculine is masculine); see §7.3 for complex controllers.

- (44) Italian (Indo-European; Romance; constructed, with inspiration of text examples)

Queste praline e questi biscotti li ho
 this.F.PL praline(F).PL and this.M.PL biscuits(M).PL 3M.PL.ACC have.PRS.1SG
dovuti letteralmente chiudere sotto chiave.
 must.PTCP.PST.M.PL literal.ADV close.INF under key(M).SG

‘Those pralines and biscuits I had to keep literally under lock and key.’

The agreement patterns in (44) can be described as an INFORMATION TRANSFER CHAIN (see Wälchli 2018) consisting of at least ten steps: The masculine plural marker *-i* (i) on the target word *dovuti* (ii) is part of the target group *dovuti chiudere* (iii), which receives masculine plural agreement from the pronoun *li* (iv), serving as controller under the condition that it precedes the verb (cf. *ho dovuto chiuder-li*, when the object clitic follows the verbal complex) (v), and is itself a target controlled by the controller group *queste praline e questi biscotti* (vi), whose gender results from gender resolution between the word-forms *praline* and *biscotti* (vii), whose word-form values feminine plural and masculine plural (viii) result from number inflection of the feminine lexeme *pralina* and the masculine

noun *biscotto* (xi), which receive their lexical gender by formal gender assignment (x).

Above we have said that the specific relationship can be of different kinds. This is the topic of the next subsection.

7.2 Specific relationships in agreement

It is not the purpose of this section to give an exhaustive treatment of all possible specific relationships that can hold in agreement. What we want to point out here is that coreference is not the only kind of specific relationship that can hold in agreement and that a specific relationship can be semantic (in that case agreement can be inter-sentential) or formal (in that case agreement is usually intra-sentential). The latter part of this section will be devoted to adjacency, which is an under-researched phenomenon often encountered in agreement. We will argue that adjacency may qualify as a specific relationship in agreement.

The clearest case of agreement blatantly violating Moravcsik's (1978: 363) Coreferentiality Principle is gender in Archi (Nakh-Daghestanian, Lezgian). Archi has excessive agreement in the clause with the absolutive argument as controller. In (45) not only the verb 'make' agrees with the absolutive argument of the clause, but the pronominal arguments in the ergative and dative cases and the adverb 'quickly' do so too.

(45) Archi (Nakh-Daghestanian, Lezgian; Bond et al. 2016: 3)

Nena<*b*>*u* *do*:[°]*zu-b* *χ'on*
 1PL.INCL.ERG<III.SG> be.big.ATTR-III.SG cow(III)[SG.ABS]
b-ela<*b*>*u* *dit*:*a*<*b*>*u* *χir* *a*<*b*>*u*
 III.SG-1PL.INCL-DAT<III.SG> quickly<III.SG> behind <III.SG>make.PFV
 'We quickly drove the big cow to us (home).'

In our view, agreement here marks the whole clause as a formal group; put differently, the agreement target is the whole clause (see §7.4 for complex targets). Agreement with the same noun class is realized wherever it can be morphologically marked in the clause. This is actually less excessive in Archi than (45) suggests, because agreement can be spelled out only occasionally in the Archi clause. Agreement appears only in about one third of the verbs, in the ergative only in the inclusive plural, in the dative only in first person pronouns, and only in 13 of 392 adverbs (Bond et al. 2016: 70). In clauses with two verbs with different absolutes, so-called biabsolutive constructions, there are two formal groups for agreement (Chumakina & Bond 2016: 90–111).

The agreement relation in Archi is a specific relationship in the sense that it is a unique relationship between the head of the absolutive NP (controller) and its clause (target). Since this is a syntactic dependency relationship, Archi agreement is intra-sentential, as opposed to coreference, which is semantic and can be inter-sentential. However, coreference is not the only possible kind of specific semantic relationship in agreement.

CO-CONCEPTUALITY – where controller and target express identity of concept, but not identity of reference – is another important specific relationship in gender agreement. Unlike co-referentiality, there is usually no agreement in number, since number is a property of the referent, not of the lexical noun expressing the concept (except for pluralia tantum, see Wälchli 2017 for Latvian). If we consider examples from the literature on anaphoric pronouns without explicit antecedents, such as *Either no letter was sent, or it got lost* and *Watch out for that snake. They are poisonous* (Bosch 1988: 211), there is no relationship of coreference between noun and pronoun. The anaphoric pronoun simply stands for something that is of the same kind as the noun (the same concept); *it* is a letter and *they* are snakes, irrespective of their reference and whether they are referential at all. The so-called “donkey sentences” also sort here. This term was first introduced by the medieval philosopher Walter Burleigh around 1328 based on the Latin *Omnis homo habens asinum* [donkey.ACC.SG.M] *videt illum* [DEM.ACC.SG.M] ‘Every man having a donkey sees it’ (Seuren 2009: 269). In languages with lexical gender, such as Latin, there is usually agreement in gender in such cases of co-conceptuality.

Co-conceptuality is particularly important for independent adjectives and numerals and other independent elements as was discussed in §4. Like anaphoric pronouns, independent adjectives typically express some sort of anaphoric relationship, which, however, does not imply identity of reference, but identity of concept. This is illustrated in (46) from German.

- (46) German (Indo-European, Germanic): independent adjective expressing co-conceptuality

Das mit dem Hemd [shirt(N)] leuchtet mir so langsam auch ein... ja, ein weißes [INDF.NOM.SG.N white.NOM.SG.N] *wäre in der Tat besser gewesen.*
[After a non-successful job application:] ‘The thing with the shirt starts becoming clear to me, too...yes, a white one would indeed have been better.’

<http://www.bewerbung-forum.de> [2018-11-06]

It might be objected that (46) is a case of ellipsis of the head noun. However, the form of attributive adjectives or numerals with head nouns and of independent adjectives or numerals without head nouns is not always the same in all languages, which is an argument that independent adjectives and numerals are not attributive adjectives and numerals with ellipsis. In German, the independent numeral ‘one’ follows a different declension pattern (originally pronominal endings): (speaking of shirts) *ein-es ist hellgrau* [one-NOM.SG.N.PRON be.PRS.3SG light.gray] ‘one is light gray’ as opposed to *ein Hemd ist hellgrau* [one.NOM.SG.N shirt be.PRS.3SG light.gray] ‘one shirt is light gray’. Wälchli (2017) discusses the case of Dundaga Latvian, where there is gender agreement only in NPs with head nouns, but not in independent adjectives (which only have number agreement; see example (31) in §4.4). As far as the Animacy Hierarchy is concerned, independent adjectives behave like pronouns. In (47) from German there is semantic (referent-based) agreement with independent adjectives rather than lexical agreement. With attributive adjectives, semantic agreement is ungrammatical (*das[N] ältere Mädchen(N)*, **die[F] ältere Mädchen(N)* ‘the older girl’).

- (47) German (Indo-European, Germanic): co-conceptuality linked to coreference by means of part-whole relationship
Zwei Mädchen [girl(N).PL] *im Alter von sieben und acht Jahren sind am Samstag in Schwarzenberg am Böhmerwald (Bezirk Rohrbach) von der Holzleiter eines Hochstandes gestürzt. Die* [DEF.NOM.SG.F] *ältere der beiden war ausgerutscht und hatte die* [DEF.ACC.SG.F] *jüngere mitgerissen.*
 ‘Two girls aged seven and eight years fell from the wooden ladder of a tree stand in Schwarzenberg am Böhmerwald (district of Rohrbach) on Saturday. **The elder one** of the two had slipped and had dragged **the younger one** with her.’
<http://www.salzburg.com/nachrichten/oesterreich/chronik/sn/artikel/zwei-maedchen-in-ooe-von-hochstand-gestuerzt-und-verletzt-209318>
 [accessed 2017-06-05]

In the rest of this section we will focus on adjacency as a further potential specific relationship in agreement.

ADJACENCY in agreement means that controller and target or target and controller immediately follow each other. Target and/or controller can consist of several words. Adjacency between controller and target is frequent in most languages with gender, which is natural, since agreement is often local. According to Corbett (2006), local agreement is more canonical than distal agreement. But most treatments of gender do not pay any particular attention to adjacency. We

think that adjacency is an important issue in agreement that deserves particular attention because there are several languages where gender agreement is predominantly or exclusively adjacent.

Since adjacency is unequivocal, it has the potential of qualifying as a specific relationship between controller and target. Thus, it is a candidate for a type of specific relationship between controller and target on a par with coreference and other specific relationships.

It is well-known that linearity plays an important role in phonology, notably in sandhi phenomena, sound changes that take place at word- or morpheme-boundaries. As already mentioned in §6.4, some instances of gender agreement originate in sandhi. This adds a developmental perspective to the study of adjacency in gender agreement. In cases where sandhi is involved in the origin of gender agreement, adjacency may reflect preservation of an earlier phonological motivation.

The importance of sandhi phenomena in agreement is well-known from Celtic languages. In all Celtic languages, feminine nouns have “mutated” onsets following an article: Irish *bean* ‘woman(F)’, *an bhean* ‘the woman’; Welsh *pont* ‘bridge(F)’, *y bont* ‘the bridge’ (Fife & King 1986: 480). Here it looks as if the controller noun is at the same time the target. However, morpho-syntactically it is rather the article which is the target with the gender marker being realized phonologically on the following word. In the case of postposed adjectives, it is just the other way round. In Breton *ur verc’h vras* ‘a big girl’ (*merc’h* ‘girl’, *bras* ‘big’), *vras* ‘big’ looks as if it displays agreement with its initial mutation, but the mutation is in fact diachronically caused by the feminine noun preceding it (Fife & King 1986: 480). (For the possessive pronoun in Welsh see Wälchli 2019 [this volume].)

Let us now turn to the discussion of languages in which target and controller in gender agreement almost always are adjacent. In Uduk, a language with two noun classes termed class 1 and class 2, gender targets immediately precede gender controllers (see Killian 2019 [in Volume I]). At the same time, coreference does not seem to play any major role. If there are two or more words in a noun phrase, the head noun and the modifier have genders of their own.¹⁸ Gender in Uduk does not usually have the function of signaling that two words belong to the same constituent or are coreferential. The gender marker is simply triggered by the gender of the following word. In (48) the preposition *kí* is followed by a class 1 noun (*yil* ‘year’). With a following class 2 noun it would be *ká*. The modifier ‘small’, however, is class 2, as are all modifiers derived from stative verbs

¹⁸For one single exception involving prenominal modifiers, see Killian (2015: 128).

with the suffix *-gàʔ*. This is why the associative marker, which links words in the NP, takes a class 2 marker. Since the gender markers are clitics in some case forms (see Killian 2019 [in Volume I]), the words on which the gender marker may appear make sometimes rather unexpected targets, such as the adverbial subordinator *gòm* in (48).

- (48) Uduk (Koman; Killian 2015: 382)
gòm=à 'cí yǐsà 'bór-ó'd á'dī kí mǎsh kī-Ø
 for=CL2 'child(CL2) NEG good:IPFV-3SG 3SG NARR marry with-CL1
yìl=à gwǎ'd-gàʔ
 year(CL1)=ASS.CL2 small-NMLZ(CL2)
 'Because it's not good for the child to marry early.'

It is not entirely obvious what the target is in this case. In one possible analysis, *gòm* is the gender target, because this word bears the gender marker. In another possible analysis, which we prefer, the gender target is the clitic *=à* which happens to require a host for phonological rather than morphosyntactic reasons.

While almost all gender agreement in Uduk is adjacent, the Taa languages West !Xõo [=West Xoon] and East !Xõo [=East Taa] (Tuul [=Southern Khoisan], Taa) have several different kinds of agreement, only one of which exhibits adjacency. Agreement preceding the controller is necessarily adjacent, agreement within the NP with the NP head noun as controller is not. Adjacency agreement is illustrated in (49) with a compound. Its first part, which is not the head of the compound, triggers agreement on the preceding word. In Taa languages, the gender of the whole compound often differs from the gender of the parts of the compound. In (49), *ʔkx'oe nʔaen* [rain house.PL] 'clouds' has gender CL2A[SG]/CL2A[PL], but *ʔkx'oe* 'rain' (only singular) has gender 3 and *nʔahe* SG (*nʔaen* PL) 'house' has gender CL3[SG]/CL1[PL]. Adjacent gender in Taa languages is always controlled by the immediately following noun, which is the first part of the compound, *ʔkx'oe* 'rain(CL3)' in (49), rather than the whole compound 'cloud(CL2A)'. In (50), there is an associative plural formed from a person name. The associative plural has gender 4. However, the adjacency agreement is triggered by the gender of the person name, which is gender 1.

- (49) West !Xõo (Tuul, Taa; Güldemann 2004): adjacency and compound gender
n si nʔa=e ʔkx'oe nʔaen ka ʔari
 1SG IPFV see=CL3 [rain(CL3) house.PL(CL1)](CL2A) REL.CL2A many
ka
 REL.CL2A
 'I see many clouds.'

- (50) West !Xõo (Tuu, Taa; Güldemann 2006): adjacency and associative plural
nna n/a=i Tom-tu ku lai k=i
 1SG:PRF see=CL1 [Tom(CL1)-ASS.PL](CL4) REL.CL4 stay OBLIQUE=CL1
dertien ku
 TOPONYM(CL1) REL.CL4
 ‘I have seen Tom and them who were at post 13.’

Some clitic hosts, such as the question particle /*V* in (51) (*V* means that the vowel must come from agreement), never occur without a following gender clitic. There is thus no gender marking, if there is no clitic host. If (51) were not a question, there would not be any gender marker.

- (51) East !Xõo (Tuu, Taa; Traill 1994: 18)
/=ú tûu à sil
 Q=CL4 people(CL4) TENSE come
 ‘Did the people come?’

A third language where adjacency plays an important role is Nalca. Gender is triggered by the immediately preceding constituent. In (52) the noun *heik* ‘hamlet’ is followed by two case markers, dative plus comitative, which together express the notion of source. Case markers and gender markers are mutually dependent on each other and hence almost always co-occur in the case-number word following the noun. In (52), *heik* ‘hamlet’ is default noun *e*-gender, which is why the first case-number word in (52) is default noun *e*-gender. However, since the controller cannot control anything but the gender of the immediately following target, the second case-gender word is default phrase *a*-gender (which is never triggered by a lexical noun). There is no adjacency condition on the demonstrative suffix, which is repeated in both case-number words.

- (52) Nalca (Nuclear Trans-New Guinea, Mek; New Testament 40021001; Wälchli 2018)
heik e-nye-k a-nye-b dara
 hamlet DN-DEM-DAT DP-DEM-COM/ABL TOP
 ‘from this hamlet’

As argued by Wälchli (2018), gender in Nalca partly derives from sandhi phenomena, which motivates adjacency diachronically.

In both Nalca and Uduk, person names are important gender controllers (see Killian 2019 [in Volume I] for Uduk). In the Oceanic languages spoken on the island of Makira, gender classes have developed from an extension of person

name markers, and, as discussed in §6.4, one of five classes in Owa originates from sandhi phenomena. Person names rarely have attributes. Thus it is natural that person name markers and person names are typically adjacent.

Adjacency-based gender developing from person name markers is not particularly complex when it first develops. There is only one agreement target, person name markers, and gender need not be specified in the noun lexicon, since it is organized by the animacy hierarchy (see §3). Person name markers can then travel down the animacy hierarchy, first expanding to older kinship terms and to other words typically expressing unique reference, in Makira languages also to the pronoun ‘who’. Both Mek languages and Makira languages illustrate complexification in terms of number of classes (in Mek from two in Una and Eipo to four to six in Nalca, depending on whether only classes with lexical controllers or all classes are counted, and in Makira from two in Arosi to four in Kahua and five in Owa). Uduk gender is considerably more complex in terms of gender assignment (see Killian 2019 [in Volume I]) and it is not known how the system has developed. Gender in Taa languages is the most complex among the languages discussed here and nothing is known about the origin of the system.

7.3 Complex controllers

Complex controllers, where the controller consists of more than one word, are well-known from gender resolution in coordination, but also inalienable possession, names consisting of several words, and nominalizations. They provide evidence for gender being assigned to a group of words rather than to a single word. In this section we will consider evidence from inalienable possession in Paumari, from German restaurant names and Taa nominalizations. The section also discusses Nalca, where complex controllers are pervasive.

The assumption of complex controllers is uncontroversial for gender resolution in coordination, as illustrated in (44) in §7.1. However, gender resolution is not restricted to coordination. Consider (53) from Paumari, a language with two different gender systems: masculine/feminine and *ka-* vs. non-*ka*-noun classes. In Paumari, there is gender resolution in inalienable possession in the *ka-* vs. non-*ka* gender system. “If either the possessor, or the possessed noun (or both) belong to the *ka-* class, a modifier takes the *ka-* class marking, no matter which one of the two it modifies” (Aikhenvald 2010: 240). Put differently, *ka*-/non-*ka* gender in Paumari inalienable possession is computed with formal criteria in the same way as in gender resolution in coordination. In (53) the possessor is *ka* and the possessed noun is non-*ka*. The adjective displays *ka*-agreement whether it mod-

ifies the possessor or the possessed noun.¹⁹ The possessor *bodi* ‘mouth(N.KA;F)’ also takes *ka-* because it agrees with *ojoro* ‘turtle(KA;F)’.

- (53) Paumari (Arawan, Aikhenvald 2010: 240): gender resolution in inalienable possession

ojoro ka-bodi-ni ka-karaho

turtle(KA;F) KA-mouth(N.KA;F)-3SG.F.DERIV KA-big

‘big mouth of a turtle’ or ‘mouth of a big turtle’

Further evidence for complex controllers comes from cases where gender is assigned on the level of group of words rather than on the level of words, which can hold for names consisting of several words. Plank (2015) discusses German restaurant names, which often can be neuter irrespective of the gender of the head noun.²⁰ The German lexeme *Orkan* ‘hurricane’ is masculine. However, in (54), *Orkan* is used as name for a restaurant and is neuter. (55) illustrates the same phenomenon with a name consisting of more than one word. *Oma* ‘grandma’ is feminine, but it is the whole expression *Oma Plüsch* ‘grandma Plush’ that is the restaurant name and as a restaurant name consisting of two words it is neuter.

- (54) German (Indo-European, Germanic; Angerer 2009: 132)

Hinter der wohl schmalsten Eingangstüre
behind DEF.GEN.SG.F probably narrow.SUPERL.GEN.SG.F entrance.door(F)

Regensburgs verbirgt sich das Orkan.

Regensburg.GEN hide.PRS.3SG RFL DEF.NOM.SG.N hurricane(M)

‘The Orkan is hidden behind the probably narrowest door in Regensburg.’

- (55) German (Indo-European, Germanic; tripadvisor.de [2018])

Das Oma Plüsch liegt direkt an der

DEF.NOM.SG.N grandma(F) Plüsch lie.PRS.3SG directly at DEF.DAT.SG.F

Donau.

Danube(F)

‘Oma Plüsch is located directly at the River Danube.’

¹⁹For similar phenomena in the related language Jarawara, where the *ka*/non-*ka*-gender was lost, see Dixon (2000) and §3.4.

²⁰According to Plank (2015), recategorization with default-neuter for German restaurants pragmatically indicates the distance of the name to gastronomy. Traditional names for restaurants, such as *Die Sonne* [the.F sun(F)] and *der Ratskeller* [the.M council.cellar(M)] are not neuter. Neuter gender for German restaurants is not obligatory.

Adjectives as parts of names commit restaurant names to the gender of their lexical head: *der Bayerische Bahnhof* [the.M.SG Bavarian.M/F/N.SG railway_station(M)], even if the adjective cannot inflect: *die Schweizer Grenze* [the.F.SG Swiss.ADJ border(F)]. This only holds if the adjective is part of the name. With non-restrictive adjectives neuter is possible: *das spießige Vier Jahreszeiten* [the.N.SG petty-bourgeois.M/F/N.SG four seasons.PL] (Plank 2015). To state this in more general terms, if the lexical head is already combined with a potential target before the name is completed, the noun phrase has already committed itself to a gender, thus gender assigned to the name as a whole is no longer available. The same rule holds, for instance, for names of roses, which can be default-feminine unless they contain an adjective (*die Helmut Schmidt, die Gruß an Helgoland* [the.F.SG greeting(M.SG) to Helgoland], but *der Gelbe Engel* [the.M.SG yellow angel(M)]).²¹

While complex controllers in German are limited to names, Nalca has them all over. In Nalca there is a general alternation between one of four lexical genders – masculine *be-*, feminine *ge-*, phonologically assigned CV-gender *ne-* (the controller has the structure CV or V), and default noun *e*-gender – and default phrasal gender *a-*, which is never controlled by a lexical noun. The switch is syntactically determined. Having certain modifiers (“allies”) helps the noun impose its lexical gender, having certain other modifiers (“obstacles”) conditions the phrasal default. Most nouns cannot impose their lexical gender unless they have an attribute ally that helps them impose their gender, as in (56) about boys’ initiation rites, where *me* ‘boy, child’ with lexical CV-gender *ne-* triggers *ne-* only if there is an adjective in the NP, but has default phrase *a-* if it is bare in the NP.

- (56) Nalca (Nuclear Trans-New Guinea, Mek; Binzell n.d.; Wälchli 2018: 71)

me ***a-ra*** *gelelinga* *soob-vka*
 child(cv) DP-TOP unnoticed enclose.in.netbag-CVB
bo-ba-lam-ek. *Nauba me* ***ne-ra*** *al-biyok*
 carry-go-HAB/IPFV-PST.3PL big child(cv) CV-TOP 3SG-alone
ba-lam-ok. *Mek me* ***ne-ra*** *soob-oka*
 go-IPFV-PST.3SG small child(cv) CV-TOP enclose.in.netbag-CVB

²¹Sometimes the gender of names is paradigmatically inherited – names of roses (F), apples (M), pears (F), beers (N), and wines (M) have the gender of their general noun as default. However, in case of restaurants (N), ships (F), motorcycles (F), and cars (M), the default-gender is not inherited from a general noun.

bɔ-ba-lam-ek.

carry-go-HAB/IPFV-PST.3PL

‘They carried the boy away secretly in a netbag. A big boy went by himself. A small boy they carried in a netbag.’

There is a parallel in Mopán Maya, where gender also has developed from an extension of person name markers. Gender in Mopán Maya is marked only on one target, the “gender marker” proposed to the noun or adjective+noun, which distinguishes masculine *aj* and feminine *ix* (Contini-Morava & Danziger 2018). Only a minority of nouns have gender, most nouns take the article *a* instead (which, unlike gender markers, is not compatible with possessive pronouns). Nouns that are not gendered when used in isolation may sometimes optionally have a gender marker if there is an attributive adjective. Contini-Morava & Danziger (2018: 138) give an example from a story where *a ch'o'oj=o* [ART rat=ECHO] ‘rat’ is first introduced without gender and then occurs with adjectives and gender markers as *aj noxi' ch'o'oj=o* [GM.M big rat=ECHO] and *aj tz'i' ch'o'oj=o* [GM.M small rat=ECHO] with a gender switch very similar to that in the Nalca example (56). The difference is that Nalca *a*- default phrase gender is formally integrated in the gender system and the alternation is more systematic in Nalca.

In some languages, sentential nominalizations can be gender controllers. Nalca sentential nominalizations, if not followed by a noun, can take one of three phrasal suffixes and each of the three resulting constructions without a nominal head takes another gender. Two of the three suffixes are homonymous and are distinguished only by the gender they control (Wälchli 2018): male nominalizations with suffix *-nya* (57) take masculine gender *be-* and thing-nominalizations with suffix *-nya* (58) take neuter gender *ne-* (which happens to have the same form as CV-gender in (56)).

- (57) Nalca (Nuclear Trans-New Guinea, Mek; New Testament, 44010021; Wälchli 2018: 80)

... [*ugun-da na e-le-nu-lum*]-*nya* *be-ra, na-ra*
 2PL-TOP 1SG²² search-IPFV-OBJ.1SG-PRS.2PL-NMLZ.M M-TOP 1SG-TOP
al-an ...
 3SG-DEM

‘... I am he whom you are looking for!’, lit. ‘I am he, the one [you are looking for me]’

²²In Nalca nominalizations, O is often zero marked, but ‘thing’ nominalizations tend to have a dative-marked O as in (58).

- (58) Nalca (Nuclear Trans-New Guinea, Mek; New Testament, 43006026; Wälchli 2018: 80)
 ... [ugun-da na-k e-le-nu-lum]-**nya** **ne-ne-ra** ...
 2PL-TOP 1SG-DAT search-IPFV-OBJ.1SG-PRS.2PL-NMLZ.N N-DEM-TOP
 ‘...you seek me not [because you saw signs]...’, lit. ‘this fact that [you are looking for me]’

Nalca nominalizations are morphologically marked, but there is also a semantic component, which is strengthened by the homonymy of two different morphological markers. There is no competition with lexical gender as there are no lexical heads in the construction. The sentential nominalization with its morphological marker must immediately precede the gender target (adjacency agreement, see §7.2).

A similar construction is found in East !Xõo, where, however, nominalizations can only take one gender. The nominalization suffix *-sà* can attach to the verb stem (*!qāhe-sà* [hunt-NMLZ(CL2)] ‘hunting’) or to a verb phrase, the subject of the nominalization being expressed by a possessor in a POSSESSOR /V+GENDER.MARKER POSSESSED construction. The preposition /V takes the gender of the immediately adjacent following controller. The only available controller is the nominalized verb phrase, which is a constituent without any lexical head from which the gender of the nominalization derives.

- (59) East !Xõo (Tuu, Taa; Traill 1994: 30; Güldemann 2004: 7)
ùh ñ bà ll̥-n /à /ùã /àũ ll̥nà
 CL4 TENSE ASPECT refuse-1SG POSS.CL2 hold/give.CL2 tobacco(CL2)
/nēe-sà
 to.3-NMLZ(CL2)
 ‘They disapprove of my giving him tobacco.’

To summarize, there is a diverse set of formal syntactic groups that can all function as complex controllers. These include NP-coordination (gender resolution), possessed and possessor in inalienable possession, names consisting of several words, nominalizations, and – in Nalca – any kind of noun phrase. Since compounds are also groups of words, we can also add compounds taking compound gender, as in Khasi (see §6.3), as a further type of formal groups serving as complex controllers.

7.4 Complex targets

Target groups or complex targets must be invoked whenever the agreement relation applies between the controller and a formal group of words. This is the case, for instance, if the target is a complex predicate consisting of several verbs (lexical verb and auxiliary) which share the same argument structure. This is most clearly visible if there is agreement with the object and the agreement is realized on an auxiliary as in (44) from Italian discussed in §7.1.

Haspelmath (1999) discusses the Italian data together with two languages where agreement goes with the absolutive to which we turn now: Godoberi (Nakh-Daghestanian, Andic) and Hindi and Urdu (Indo-European, Indo-Aryan). In (60) from Hindi/Urdu, the feminine noun ‘bread’ is the object of ‘eat’, but all three verbs in the verb complex display agreement. This means that the three verbs together make up one complex target. (61) from Godoberi shows so-called “long distance agreement”, which Haspelmath (1999) analyzes as an instance of clause-union. In our terms, the four verbs in (61) together constitute a formal group sharing the object argument and are a single complex target, with the neuter plural of the absolutive realized on three of them (‘want’ never takes agreement).

- (60) Hindi/Urdu (Indo-European, Indo-Aryan; Wunderlich 1994: 23; Haspelmath 1999: 147)
Raam ne roṭii khaa-nii caah-ii thii.
 Ram ERG bread(F)[SG] eat-INF.F.SG want.PST-F.SG be.PST.F.SG
 ‘Ram had wanted to eat bread.’

- (61) Godoberi (Nakh-Daghestanian, Andic; Haspelmath 1999: 143)
ilu-li quči-be r-al-u r-uL-i
 mother-DAT book(N)-PL[ABS] PL.N-read-CVB.PST PL.N-finish-INF
q°araŋ-anta ru-k’a.
 want-CVB.PRS PL.N-be-AOR
 ‘Mother wanted to finish reading the books.’

Nakh-Daghestanian languages are known for their extensive clausal agreement, which takes different forms in different languages. In Godoberi all verbs of a unified clause together constitute an agreement target (see also the similar case of Archi in §7.2).

Complex gender targets involving complex predicates also occur in Coastal Marind. In (62) the patient *ebta* ‘sago thatch’ is an argument of the transitive

verb *takun* ‘make roof’, but agreement is shown on the auxiliary *balen* ‘finish (intr./tr.)’.²³

- (62) Coastal Marind (Anim, Marindic; Bruno Olsson, p.c.)
ebta takun mbya nak-ap-ba-<h>in
 sago.thatch(IV) make.roof NEG 1.A-CONTESSIVE-finish<IV.U>
 ‘I didn’t finish making the sago thatch roofing.’

In discussing agreement in case, Lehmann (1982: 222) points out that viewing the head noun as controller of agreement is problematic notably when an NP lacks a head noun. This also holds for gender in independent headless NPs, such as Italian *Tu sei la più bella* ‘you are the most beautiful one (F)’ (see §4). Here it is obviously not the article controlling feminine gender on the adjective or vice versa, but the whole headless noun phrase in the predicate is a target group assigned feminine singular by a latent contextual controller.

If attributes in headless NPs form target groups, the question arises as to whether a series of target words within the same NP could generally be considered to constitute a target group. In many languages gender agreement with multiple targets in an NP is a way to signal that these elements all belong together in one formal group (which can be contiguous or non-contiguous). This would then mean that in NP agreement the head noun is the controller and the whole NP is the target. A potential problem is then that the head noun controlling the NP is also part of the NP. Lehmann (1982: 223) suggests that this could be solved with the following condition “If B is the head of an NP A, B is not said to agree.” This may seem entirely ad hoc at first glance. However, if we take into account that agreement is displaced information, it is a priori excluded that the controller can be part of the target. Target groups are formal groups, but not all formal groups are syntactic constituents. The easiest solution is to say that in NP agreement, the target group is the NP minus the head noun. What we have said for NPs here also applies to clauses in the Daghestanian languages Archi and Godoberi where the whole clause can be the agreement target (see §7.2).

A further case of complex targets are gendered clauses in Ngan'gityemerri serving as relative clauses, discussed in §4.3.

Table 16 summarizes the kinds of formal groups involved in complex controllers and complex targets mentioned in §7.3 and §7.4.

²³For similar phenomena involving person in another language in South New Guinea, Nen (Morehead-Wasur), see Evans (2015).

Table 16: Formal groups serving as complex controllers and complex targets in gender agreement

Formal groups manifest in gender agreement	
Complex controllers	NP coordination Inalienable possession Names consisting of several words Nominalizations Compounds Complex noun phrases
Complex targets	Complex predicates Clauses Noun phrases Gendered clauses (relative clauses)

7.5 Features as mature conditions

Morphosyntactic features are a highly mature form of information transfer. In non-mature gender systems it is often difficult to identify a [NUMBER] feature. For instance, Ngan'gityemerri has a noun class *awa-* glossed ‘mob’ for a group of people (Reid 1990: 296), but number does not otherwise interact with gender. If we consider what makes gender a good feature, it is pretty much the same characteristics that are traditionally invoked for delimiting genders from classifiers: there is a closed set of values with up to twenty members (Dixon 1982: 215), the same system of values applies to different targets, all nouns are controllers, and gender markers are bound elements on target words. All these properties are indications of maturity (see also §6). In our view, features are emergent and develop through grammaticalization, thus there is no reason to assume a universal set of morphosyntactic features. The existence of languages with two parallel concurrent gender systems, such as Paumari (Aikhenvald 2010) and Burmeso (isolate; Donohue 2001), is an argument against a universal set of features (see also Dahl 2000b, Corbett et al. 2017: 252 and Svärd 2019 [in Volume I], Liljegren 2019 [in Volume I], and Sinnemäki 2019 [this volume], for other languages with two parallel gender systems).

Further evidence that features are not all there is to displaced information in agreement comes from what Corbett 2006 calls conditions. Conditions are “factors which are not themselves realized directly in agreement” (Corbett 2006: 176).

As a rule of thumb, features, but not conditions, are usually glossed. Many examples of conditions pertain to the realm of animacy and related notions such as individuation. In Miya (Afro-Asiatic, West Chadic), attributive demonstratives take plural agreement only if the controller is animate. Since the masculine-feminine gender opposition is neutralized in the plural in Miya, this entails the peculiar pattern that in the plural masculine and feminine are realized only with inanimate controllers: *nākən víyayúw-awàw* [this.M.SG fireplace(M)-PL] ‘these fireplaces’ (Schuh 1998: 193; Corbett 2006: 178). Recall from §7.4 that groups of verbs in some Nakh-Daghestanian languages can form target groups, a phenomenon often referred to as “long distance agreement”. In Tsez (Nakh-Daghestanian, Tsezic) “long distance agreement” is conditioned by topicality. A target group of several verbs agrees with the absolutive of the subordinate verb only if the S or O of the subordinate clause is a topic (Polinsky & Comrie 1999; Corbett 2006: 197).

Conditions are conditions on agreement. As a consequence, if a condition turns into a feature, the result is usually a combination of two features in cumulative exponence. If features develop from conditions it is no coincidence that features often cumulate with each other. Since animacy is a very frequent type of condition, it is no coincidence that animate gender or other gender values reflecting animacy frequently cumulate with other agreement features, such as number (this is the topic of §8).

Corbett (2006, chap. 6) distinguishes absolute conditions, factors that always determine a certain choice of agreement value (the two examples given so far in the previous paragraph), and relative conditions, factors that favor a certain optional choice of agreement value. We change the terms to *obligatory* and *optional*, which we think are more easily understandable. In Russian, controllers consisting of two conjuncts are more likely to trigger plural agreement when animate than when inanimate, which is an instance of an optional (relative) condition on agreement (Corbett 2006: 179).

When conditions develop into features, it is reasonable to assume that they are first optional. This suggests the following grammaticalization path (63):

- (63) Grammaticalization path from condition to feature
optional condition on agreement > obligatory condition on agreement >
gender (= cumulative feature)

Consider the example of (in)animate subgenders in Russian and other Slavic languages (Corbett 1991: 42, 2006: 118; see also §3.5). In Russian, only the major declension class for feminine nouns has dedicated accusative forms, and only in the singular. Masculine singular nouns and all plural nouns take the genitive form

if animate and the nominative form if inanimate. In Serbian-Croatian-Bosnian, only the masculine singular is affected, so there are only two subgenders in the masculine. Slavic (in)animate subgenders originate as a condition on case, but in Russian animacy has gone quite a long way to become lexical gender, as the subgender of most nouns is fixed irrespective of their referent-based animacy (see §3.5). For instance, *konkurent* ‘competitor’ is always animate; however, *duši* ‘souls(F)’ (in feminines, the animacy distinction is visible only in the plural), The Pentagon and The White House are never animate. Russian has undergone the development in (63). Huntley (1980) surveys evidence from several Slavic languages demonstrating how the category was extended from object function to use with other functions of the accusative with prepositions, and from definite human to human and animate. In Polish the genitive singular form is further extended to individualized inanimates (Björn Wiemer, p.c.).

In Slavic there was already gender (masculine, feminine and neuter) before the development in (63). The path in (63) is possible also when there is no gender originally. However, there must be some form of agreement already. An interesting example in this respect is Lakhota (Siouan) with plural actor and under-goer agreement on the verb with animate nouns, which Sinnemäki (2019 [this volume]), following Van Valin (1977: 36–37), classifies as an instance of gender. Another possible interpretation is that the “enclitic =*pi* indicates plurality of all human subjects” (Mithun 1999: 508) and that there is no verbal agreement at all in Lakhota verbs. The question as to whether animacy in Lakhota can be interpreted as a feature is very much dependent on how number, which it conditions, is interpreted. A condition cannot turn into agreement if the category which it conditions is not agreement.

Pnar attributive adjectives, discussed in §4.3, illustrate that an animacy distinction can emerge in a language with gender without connection to that gender system. Recall from §4.3 (example (29)) that one type of attributive adjectives in Pnar optionally takes the proposed nominalizer *wa*. However, with human head nouns, the nominalizer *wa* is obligatory with this adjective type. This is an optional condition as far as non-human nouns are concerned, and an obligatory condition as far as human nouns are concerned.

We may conclude that features can evolve from conditions on agreement and that if there is a feature in agreement already, another one, especially if animacy-based, can more easily join it in cumulative expression (realized by the same marker).

7.6 Nominal gender targets and the decategorialization of nouns

Agreement usually has nominal controllers and non-nominal targets. Nominal is used here in the sense of a cover term for nouns, noun phrases and formal groups of nouns. However, nouns outside their prototypical discourse function of referring (Croft 2001: 87) in modification or predication use tend to lose some of their nominal properties. Hopper & Thompson (1984: 711) call this decategorialization. Decategorialization of nouns is highly relevant for gender, since the possibility to serve as a target for gender may be a property of nouns undergoing decategorialization.

An important kind of nominal target is adnominal possessors. The double nature of possessors is most obvious in independent possessors which can either agree with the possessed or with the possessor (the latter is person indexing) and in some languages, such as German (see, e.g., Wälchli 2019 [this volume]) and Biak, they agree with both.²⁴ Adjectivized possessors are more inclined to agree with the head noun than nominal possessors. However, adjectivization does not always preclude possessors from being controllers for modifiers themselves, as in (64) from Upper Sorbian. It is unexpected that the Sorbian adjective can trigger agreement in (64), but given that this is the case, it is not unexpected that gender here is referent-based (since there is no lexical noun that could trigger the agreement).

- (64) Upper Sorbian (Indo-European, Slavic; Schuster-Šewc 1976: 27; Corbett 2006: 62)
- | | | | |
|----------|------------------------|--------------------------|-----------------|
| <i>w</i> | <i>[naš-eho</i> | <i>nan]-ow-ej</i> | <i>chěž-i</i> |
| in our- | GEN.SG.M | father-POSS.ADJ-LOC.SG.F | house(F)-LOC.SG |
- ‘in our father’s house’

NOMINAL GENDER TARGETS (nouns or noun phrases that are gender targets) are a heterogeneous group of phenomena where a noun or noun phrase looks as if it was an agreement target of another noun or NP. (65) from German is an example of a nominal gender target. Most German nouns for professions have to mark gender derivationally (derivational gender). The predicate noun carries a redundant marking whose choice is determined externally, in (65) by the referent of the subject.

²⁴Biak (Austronesian, Cenderawasih Bay) distinguishes animates and inanimates only in the plural. Body parts that occur in pairs are often animate, as in *tanduk* *v<y>e=s-ya* [horn <3SG>POSS=3PL.ANIM-SPEC] ‘its horns (of one animal)’ (van den Heuvel 2006: 106). Excrements, such as ‘spit’, are plural and inanimate: *an infse=na* [NMLZ spit 3PL.ANIM.POSS=3PL.INAN.SPEC] ‘their spit (of those people)’ (van den Heuvel 2006: 273).

- (65) German (Indo-European, Germanic): predicate professional noun marked for gender

Angela Merkel ist die beste
 Angela Merkel be.PRS.3SG DEF.NOM.SG.F best.NOM.SG.WEAK
Kanzlerin, die wir je hatten.
 chancellor.DERIV.FEM REL.ACC.SG.F we.NOM ever have.PST.3PL

‘Angela Merkel is the best chancellor we ever had.’

www.plattentests.de/mobile/forum.php?action=showThread&id=89713
 [2018-10-10]

Despite its female derivational suffix *-in*, *Kanzlerin* in (65) denotes the whole set of male and female Chancellors of Germany (otherwise the set could not be restricted by ‘best’), among which there only was a single female one so far. The same holds when Margaret Thatcher in 2013 was called *Großbritanniens umstrittenste Premierministerin* ‘Great Britain’s most controversial prime minister’ (www.spiegel.de › Politik › Ausland › Tories Apr 11, 2013).

While adjectives do not agree in predicative position in German, superlative predicates mark gender agreement in the singular on the article. The superlative predicate necessitates a forced choice of gender, which is determined externally. In (66) there are two competing NPs with different lexical gender, differing also in their level of taxonomy. In German there is usually agreement by co-conceptualization with the hyperonym in the construction type instantiated in (66), in Latvian with the hyponym (67), and Italian is mixed, as illustrated in (68–69).

- (66) German (Indo-European, Germanic): agreement by co-conceptualization with hyperonym

Von allen Tieren ist der Löwe
 from all.DAT.PL animal(N).DAT.PL be.PRS.3SG DEF.NOM.SG.M lion(M)
das majestätischste.
 DEF.NOM.SG.N majestic.SUPERL

‘Among all animals the lion is the most majestic one.’

- (67) Latvian (Indo-European, Baltic): agreement by co-conceptualization with hyponym

No visiem zvēriem lapsa ir
 from all.DAT.PL.M animal(M).DAT.PL fox(F).NOM.SG be.PRS.3

visgudrākā.

all.smart.COMP.NOM.SG.F.DEF

‘Among all animals, the fox is the smartest one.’

- (68) Italian (Indo-European, Romance): agreement by co-conceptualization with hyponym

Tra tutti i fiori la rosa è

among all.PL.M DEF.PL.M flower(M).PL DEF.SG.F rose(F).SG be.PRS.3SG

la più bella.

DEF.SG.F more beautiful.SG.F

‘Among all flowers, the rose is the most beautiful one.’

- (69) Italian (Indo-European, Romance): agreement by co-conceptualization with hyperonym

Tra tutti i paesi la Svizzera

among all.PL.M DEF.PL.M country(M).PL DEF.SG.F Switzerland(F).SG

è il più neutrale.

be.PRS.3SG DEF.SG.M more neutral.SG

‘Among all countries, Switzerland is the most neutral one.’

Nominal targets are highly relevant for gender from a diachronic point of view since it is well-known that gender markers can grammaticalize from nouns (Heine & Reh 1984: 225). Since grammaticalization from nouns to gender markers is gradual, there must be intermediate cases between noun targets and agreement proper with non-noun targets.

Yagua (Peba-Yagua) and other Amazonian languages demonstrate how agreement with noun targets can gradually give rise to agreement by decategorialization of nouns. Yagua has a large set of classificatory formatives, many of which can be shown to originate from nouns (Payne 1986: 120), such as *jaá* ‘water’ which is also the classifier for liquid. In attributive constructions as in (70), the classifier can be repeated, which looks like agreement.

- (70) Yagua (Peba-Yagua; Payne 1986: 126)

jityaá-jaá vánuqui-jaá

breast-CLF:LIQUID hot-CLF:LIQUID

‘hot milk’

Based on evidence from another Amazonian language, Miraña, which is not genealogically related to Yagua, Grinevald & Seifart (2004: 278–279) argue that

noun classes may grammaticalize from such constructions as (70) in Yagua where classifiers are used as repeaters.

It is particularly interesting in Yagua that different kinds of elements display different degrees of decategorialization. Attributes expressing qualities in Yagua are nouns and not adjectives and can also carry a non-classifying nominalizer, as in *mucata-y-sara* [boil-INTR-NMLZ] ‘boiled’. The major function of the classifier in modifiers is to nominalize the modifier and marking is actually rare, since many adjective-like concepts are inherently nominal and need not be nominalized (Payne 1986: 127). However, with demonstratives and numerals, decategorialization is more advanced. Demonstrative and numeral roots cannot stand without suffixation of a classifier, but classifiers do not cause a change in word class (Payne 1986: 127). Agreement is not obligatory, since the general inanimate classifier *-ra* can be used on a demonstrative with any head noun.

When nominal targets develop into agreement proper, the agreement marker may originate from a noun, as in Yagua, but it can also originate from a nominal derivation marker. Dressler & Doleschal (1990) show that Italian agent nouns in appositive use, such as *una risposta rivelatrice* [one.SG.F answer(F).SG reveal.AGN.F.SG] ‘a revealing answer’, *uno sguardo rivelatore* [one.SG.M glance(M).SG reveal.AGN.M.SG] ‘a revealing look’ agree in gender, which testifies to their adjectivization (see also Luraghi 2015: 75–76 for examples from other Indo-European languages).

A development from nominal targets to agreement proper also occurs in cases of gendered clauses turning into relative clauses, as in Ngan'gityemerri (Reid 1997) discussed in §4.3.

Decategorialization of nouns also occurs in the development of person name markers as in Iraya (Austronesian, North Mangyan; data from the New Testament) *laki Howan* ‘John’ (from *lalaki* ‘man’) and *bayi Mariya* ‘Mary’ (from *babayi* ‘woman’), *laki Satanas* ‘the devil’. For the development of nouns and NPs to anaphoric gender markers see Wälchli (2019 [this volume]). As shown by Mithun (1986), object noun incorporation may develop into a marker of verb classification. In the Northern Iroquoian languages, the incorporated elements are nominal, as in (71):

- (71) Cayuga (Iroquoian, Northern Iroquoian; Mithun 1986): noun incorporation in classificatory use
So:wá:s akh-náhskw-aę’
 dog I-domestic.animal-have
 ‘I have a (pet) dog.’

In the Southern Iroquoian language Cherokee, only relics of noun incorporation are left in the form of distinctions of a closed set of choices for a few verbs ('to give a living thing/liquid/a long, rigid object/a flexible object/else') (Mithun 1986: 392). According to Mithun (1986), verb classifiers may express noun classification. Passer (2016a), however, emphasizes the differences between (supposed) verb classifiers and nominal classification based on a diverse sample of thirteen languages. Even though it is a matter of debate how far verb classifiers can reach in becoming classifiers, they certainly belong to the complex of phenomena where decategorialization of nouns is involved in the development of some sort of asymmetric coreference relationship, even though it is not the core function of verb classifiers to classify nouns.

7.7 Target-controlled gender and idiomatization of gender agreement

The basic idea of the notion of agreement is that the feature value is selected by the controller. However, in some cases, the target contributes to the choice or selects the value entirely, which, similarly to nominal targets treated in §7.6, contributes to make agreement fuzzy.

Mohawk (Iroquoian) has four genders: masculine, feminine-indefinite, feminine-zoic, and neuter. Neuter differs from feminine-zoic only by not allowing for dual and plural number. Gender is expressed cumulatively with number and person in verbal prefixes. According to Mithun (2014: 155), relatively few verb stems can be used with either animate or inanimate arguments. "[V]erbs for growing, catching, burying, and having a proper name require grammatically animate patients, that is, they routinely occur with Zoic Patient prefixes" (Mithun 2014: 155). The verb for getting ripe, however, requires neuter gender. The gender for corn, for instance, is zoic when it is described as growing or short and neuter when it is ripe or dry (72):

(72) Mohawk (Iroquoian, Northern Iroquoian; Mithun 2014: 154)

o-nenhst-E' *ken'=ok*

N-CORN-NOUN.SUFFIX small=just

ni-konti-hneni-es-on's

PARTITIVE-3ZOIC.PL.AGT.length-be.long-DISTR

'The corn (i.e., corn stalks) are (ZOIC) very short.'

In Mawng, there are five genders, masculine, feminine, land, vegetation, and edible, which, among other things, are distinguished for S and O arguments in verbal prefixes (A arguments distinguish only masculine vs. non-masculine)

(Singer 2012: 984). However, many verbs tend to have different meanings with different gender prefixes. At the same time there are few overt nouns (Singer 2018: 117). Each gender has several semantic domains associated with it. For instance, liquids are land gender, plant food is edible gender, most animals are masculine, and crabs are feminine. Hence, the Mawng verb *wa* ‘consume’ usually means ‘drink’ with land gender, ‘eat plant food’ with edible gender, ‘eat animal food’ with masculine gender, and ‘eat crab’ with feminine gender. In other instances, gender marking on verbs is even more idiomatic. For instance, the Mawng verb *-apti* ‘have, hold’ tends to have land gender when used in the meaning ‘understand’. Explicit objects are often missing and most nouns for knowledge are land gender, but *mayali* ‘knowledge’ in (73) is vegetation gender. With this noun, *-apti* ‘understand’ can either take controller-induced vegetation gender or target-induced land gender:

- (73) Mawng (Iwajdian Proper; Singer 2012: 972)
K-ang-apti-Ø *ma-lijap*
 PRS-3N_M>3LAND-understand-N_PST VEGETATION-little
mayali.
 knowledge(VEGETATION)
 ‘She understands a little bit of knowledge.’

When asked to express ‘drink blood’ with the noun *maningul* ‘blood (vegetation gender)’, a native speaker prefers target-induced land gender (Singer 2012: 970), since liquids are usually land-gender.

Controller-induced gender is nothing else but lexical gender (see §3). Target-induced gender is the verbal equivalent of referent-based gender. Target-induced gender and referent-based gender are both opposed to lexical gender. If the term verbal gender were not already taken (*genus verbi* = voice), we might use this label here for the classification of events rather than referents. Singer (2012: 978) draws the parallel to classificatory noun incorporation in Mawng’s neighbor Bininj Kun-Wok (Gunwinyguan) (see §7.6 for noun incorporation in Iroquoian).

Mawng also has many cases of so-called lexicalized agreement (agreement with an argument that does not exist; Singer 2011). For instance, the verb *-marranyi* ‘wave (at OBL)’ always has third person land gender in the prefix where direct object is marked, but never has an identifiable direct object. According to Singer (2011: 640) lexicalized agreement is also found in a number of other Northern Australian languages spoken near to Mawng, such as Tiwi (isolate) and Gaagudju (isolate). It also occurs in Southern Tiwa (Kiowa-Tanoan; Frantz 1995: 84, “empty arguments”) and in Ket (Yeniseian). However, Ket pseudo-actant

markers (Vajda 2003: 79) in, among other things, involuntary causatives and stative resultatives, differ from Mawng in that they always can be interpreted as (default) neuter gender. Despite the complexity of the Ket verb morphology, this is actually not that much different from dummy subjects in Germanic languages such as English *it rains*.

If we extend the notion of lexicalized agreement to free pronouns, idioms with pronouns such as English *to make it* ‘to succeed’ or *to rough it* ‘to live without usual conveniences’ (famous through Mark Twain’s travel book *Roughing It*) can also be considered idiomatized agreement. An example with a masculine idiomatized pronoun from a Germanic language variety is Bernese Swiss German *er git ihm!* [he give.PRS.3SG him] ‘he makes an effort, hurries up’ (Greyerz & Bietenhard 1997: 125) with a semantic shift ‘hit a male person in a fight’ > ‘make an effort’. An example of a gender relic in an idiom in Germanic is the specification of time in more conservative varieties of Standard Swedish with the feminine personal pronoun *hon* ‘she’:

- (74) Swedish (Indo-European, Germanic; Teleman et al. 1999: 276): feminine gender relic with time idiom
Hur mycket är klockan/hon? – Hon är väl bortåt tre.
 how much be.PRS clock.DEF.SG.CM/she she be.PRS well towards three
 ‘How much is the time/“she”? – “She” is around three, I guess.’

Idiomatization involving gender agreement may take many different shapes. In the Torricelli language Walman (Dryer 2019 [in Volume I]), masculine is mainly restricted to human males, some larger animals and a few quasi-animate natural phenomena. In a few idioms, however, nouns that are usually feminine or pluralia tantum are masculine, notably *olokol* ‘mountain(PLT)’ and *anako* ‘sky(F)’ in idioms for ‘to thunder’ and *won* ‘chest(F)’ in idioms expressing emotions.

If gender is only retained in idioms, it disappears as a grammatical category. In this, gender is not different from any other grammatical category. In Iwaidja (Iwaidjan Proper), which is related to Mawng, gender is lost entirely and in Garig-Ilgar (Iwaidjan Proper), it is reduced to a two-value system (masculine vs. non-masculine) (Evans 2000: 115). Relics of object gender agreement can only be found in idioms (Evans 2000 calls them “pseudo-argument affixes”). Neuter gender (=Mawng land gender) and vegetable gender in Garig-Ilgar and Iwaidja still appear with a few verb roots, such as ‘consume’ and ‘know’ in idiomatic expressions in contexts where it is productive in Mawng (Evans 2000: 116; Singer 2011: 643). This can be compared to the many idioms in Swedish that retain case endings, as *till handa* ‘at hand’ and many other examples with an old genitive plural ending *-a*.

7.8 Summary

Agreement is prototypically a relationship between nouns and noun-associated forms. The prototypical discourse function of nouns is to express referents and nouns have a tendency to decategorialize if they are used in other functions, such as predication and modification. Decategorializing nouns and noun phrases gradually lose their ability to refer by themselves and some of their marking can then be reanalyzed as displaced information of referring expressions elsewhere in discourse. This displacement of information need not be syntactic, but can also be paradigmatic. There is not always an overt controller, which makes it impossible to view agreement as a purely syntactic process.

In this section we have seen that agreement is much more complex than just a syntactic relationship between two words. The relationship can be semantic and agreement can be inter-sentential. Both controllers and targets may be complex and consist of several words. To the extent that agreement is syntactic, its function is to indicate formal groups, and these formal groups can be of three different kinds: controller groups, target groups and the grouping of controller and target. Even though agreement has the potential of indicating discontinuous groups with considerable distance between the elements, agreement is often local and it is not uncommon for controller and target to be adjacent. In several cases from widely different languages, gender agreement requires adjacency, which is an underresearched phenomenon. Much of the fuzziness of agreement derives from the fuzziness of coreference, the most important specific relationship that can hold between controller and target. However, as we have seen in §7.2, coreference is by far not the only kind of relationship between controller and target.

8 Cumulation of gender with number, case and person

Gender marking systems are more often than not conflated with the encoding of other morphosyntactic features such as number, case, and person. In §8.1, we consider cumulation with number, in §8.2 cumulation with case and/or person. §8.3 puts cumulation into the wider context of the formalization of gender.

8.1 Gender and number

Patterns of interaction between gender and number seem to be particularly prominent in the functioning of gender systems, and, in fact, number is claimed to be “the category most often realized together with gender” (Corbett 1991: 189). Creissels et al. (2008) formulate an Africa-specific generalization on the nature

of this relation. They claim that African languages devoid of gender tend to have less grammaticalized strategies for the marking of nominal plurality, whereas in languages with gender, number distinctions tend to be obligatory and expressed both through nominal and non-nominal marking, often in cumulation with gender. Di Garbo (2014) and Di Garbo & Agbetsoamedo (2018) bring empirical support to this claim by investigating patterns of exponence of gender and number values in two partially overlapping samples.²⁵ Di Garbo (2014) is based on a sample of 100 African languages (84 with gender, 36 without). The sample used by Di Garbo & Verkerk (2018) is based on the gendered subset of the dataset in Di Garbo (2014), and thus consists of 84 languages, all of which have gender. In line with Creissels et al. (2008), the study by Di Garbo (2014: 134) reveals that, in the languages of Africa, pervasive patterns of encoding on noun-associated forms almost always involve both gender and number, and that, in the absence of gender, number marking tends to remain optional and to operate at the phrasal level (one marker per noun phrase). The study also concludes that cumulative exponence of gender and number is by and large the most pervasive pattern of encoding in both nominal and non-nominal (noun-associated forms) domains of gender marking. Out of a sample of 84 languages, only the North-Central Atlantic language Wamey is found to display non-cumulative encodings of gender and number, both on nouns and on all relevant noun-associated forms. In this language, however, non-cumulative exponence of gender and number is the result of a recent innovation whereby the plural prefix of gender 1/2 (to which human nouns are typically assigned) became the default plural marker, generalized to all nouns and gender- and number-inflecting forms, independently of the animacy of the noun referent (Di Garbo & Agbetsoamedo 2018: 187). A similar development is attested in the Kinshasa variety of Lingala (Atlantic-Congo, Central-Western Bantu), but only in the nominal domain. In Kinshasa Lingala, nouns can receive double plural marking: by means of a cumulative gender/number marker and the plural prefix *ba-*, which, as in the case of Wamey, originally was the plural prefix for nouns of gender 1/2, most typically human, but which is now used as a generalized plural marker, with human and non-human nouns alike (Di Garbo & Agbetsoamedo 2018: 188). In addition to investigating the distribution of cumulative exponence of gender and number, Di Garbo & Agbetsoamedo (2018) also survey the occurrence of gender syncretism in the context of non-singular number values. The results show that syncretism of gender in the context of number is also very widespread in the languages of the sample (attested in 67 out of 84

²⁵See also Güldemann & Fiedler (2019 [in Volume I]) for a thorough discussion of co-exponence of gender and number in Niger-Congo gender systems.

languages), and that its occurrence always presupposes cumulative exponence of gender and number values.

These findings offer an interesting parallel to earlier results by Carstairs (1987), who finds a similar relationship between syncretism and cumulative exponence in the domain of case and number marking: case distinctions are more likely to be syncretized in the context of plural number than any number value in the context of any case distinction. In addition, these patterns of syncretism always presuppose cumulative exponence between the two features. Carstairs (1987) interprets these findings as pointing to the existence of functional asymmetries between case and number. Di Garbo & Agbetsoamedo (2018: 205–206) suggest that the same reading could be applied to the results on gender and number. When non-cumulative exponence of gender and number emerges from the reanalysis of earlier cumulative systems of encodings (as in the case of Wamey and Kinshasa Lingala), this is likely to be linked to the development of new (and initially semantically motivated) strategies for the marking of nominal plurality. Similarly, the distribution of patterns of syncretism involving gender and number is strongly asymmetrical, with gender – and not number – being the morphosyntactic feature that is most likely to be syncretized.

There are various ways in which an asymmetric relationship between gender and number makes sense from a functional point of view. On the one hand, number has a more obviously semantic core function than gender. On the other hand, if gender preferably tends to develop in markers that already express another grammatical category, then the functional asymmetry between gender and number must also be interpreted in a developmental perspective. This is well in line with Nichols' (1992: 142) hypothesis that “agreement triggers noun classification (rather than vice versa)”. Here are some examples where it has been argued that gender markers have developed in close connection to number markers.

In various Berber and Semitic languages, the feminine *t* also has singulative and diminutive functions. Mettouchi (2000: 221) argues that the diminutive and singulative (partitive) function of the *t*-marker in Berber is diachronically prior to the feminine function (see also §5.2). Similarly, it has been suggested that the Arabic gender system was not sex-based originally. Moscati (1964: 86) speaks of “a more complex system of classes within which the category of number has to be included as well”.

The Khasian languages have innovated feminine pronouns for second and third person singular (Daladier 2011: 184), which at least partly seem to derive from the second and third plural forms not distinguishing gender with different vocalism for singular and plural forms (Khasi 2PL *phi*, 2SG.F *pha*, [vs. 2SG.M *me*], 3PL *ki*, 3SG.F *ka*, [vs. 3SG.M 'u]).

Interesting is also the case of Yagua mentioned by Wälchli (2019 [this volume]), where a woman who has given birth to a child or children is referred to with dual number. Payne (1985: 42) does not consider Yagua to have gender, but Yagua is obviously an example of a language where sex can condition the use of number.

8.2 Gender and case and person

Given the pervasiveness of number as the the category type most obviously connected with gender, any other category type will look meager in comparison. Moreover, case cannot be expected to be equally prominent because case is more restricted cross-linguistically than number. However, we think that case is also very relevant for the cumulative character of gender and this mainly for two reasons.

First, gender in anaphoric function in free and bound pronouns tends to exhibit some form of suppletion or neutralization according to grammatical relation (that is, grammatical case, if case is not restricted to dependent marking, but also includes indexical head marking on verbs), as shown by Wälchli (2019 [this volume]) specifically for feminine gender (but there is no reason to believe that feminine is exceptional in this respect). In this function, case occurs together with gender most typically in personal pronouns and pronominal affixes. Hence, here we deal with cumulation of gender with person and case rather than with case only. In addition, one person, the third, is clearly more dominant than others, and, within third person, the third person singular is more dominant than the plural, which in turn brings us back to the dominance of number as the feature with which gender interacts the most.

Second, there are several instances where gender displays systematic syncretism patterns with case, which can sometimes be shown to go back to the very origin of gender. In other instances, the origins of the patterns remain unexplained.

A well-known source of animacy in gender is differential case marking. In §3 and §7.5 we have discussed the example of Slavic, where animacy in gender has developed from differential object marking. Luraghi (2011: 456) argues that the neuter vs. non-neuter distinction in Indo-European has developed from differential subject case marking. In both Slavic and Proto-Indo-European, the origin of gender from case entails a cumulation of gender and case marking, with case in actor and undergoer roles neutralized in the less animate gender. In both Slavic and Proto-Indo-European, there is already case agreement within the NP when gender develops. In Indo-European, forms from two different demonstrative stems, animate (**so*) and inanimate (**to*), were integrated into an already existing case agreement system (Luraghi 2011: 456).

Two instances where the origin of pervasive syncretism patterns between case and gender are not known are Algonquian and Uduk. Algonquian languages have systematic syncretism between singular obviative and inanimate plural (where proximate and obviative are not distinguished; see Table 15 in §6.4). In Uduk, there is a syncretism of class 1 ergative case and class 2 accusative and associative cases (see §5.2 and Killian 2019 [in Volume I]).

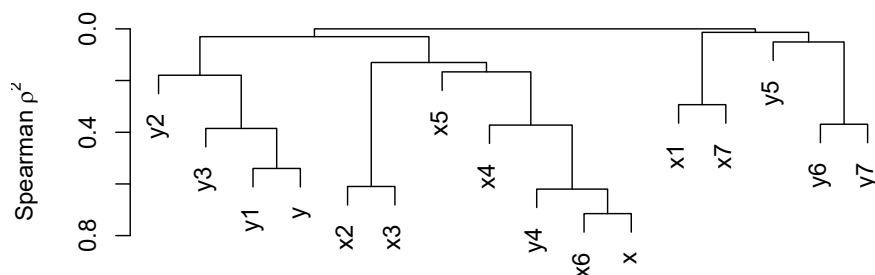
In some languages of New Guinea, notably in Nalca (Mek) and in Abau (Sepik), gender and case are expressed in the same word adjacent to the head noun. Svärd (2019 [in Volume I]) speaks of “case marker hosts”. In Mek, it can be shown that gender was originally restricted to a few postpositions distinguishing case functions and was secondarily extended to other postpositions in Nalca by analogy (Wälchli 2018).

What links gender together with case in several of the instances discussed so far is animacy (see also §3 and §3.2). While connections between gender and case due to animacy effects can be expected to be related predominantly to grammatical case, there are also interesting connections between gender and local cases. In some languages, locatives are well-connected with gender systems, in others they are completely outside of it. In many Bantu languages, locatives are integrated in gender systems (see, e.g., Bresnan & Kanerva 1989 for Nyanja [=Chicheŵa]). In Meskwaki, however, the locative case lacks gender or number distinctions (Thomason 2003: 12). In several of the Oceanic languages spoken on the Island of Makira, a place gender is developing from the local preposition *i* (see §6.4). These languages thus can help us understand how locative and gender can be intertwined. In Owa, *i* can still be interpreted as a preposition when used in isolation, but in the “sentence medial” form, used among other things before objects and following prepositional verbs, nouns of the location class (mainly place names) must take *ki* (<*k+i*), which is *k*- plus class marker: *tanga-a k-i Jerusalem* [to-3SG MEDIAL-LOC J.] ‘to Jerusalem’ as opposed to *tanga-a k-o Herod* [to-3SG MEDIAL-M H.] ‘to Herod’. Therefore, Mellow (2013: 26) lists zero for “sentence initial” and *ki* for “sentence medial” article forms of the location-noun class.

8.3 Cumulation and the degree of formalization of gender

In this section we will argue that there is a correlation between cumulation of gender with other grammatical categories and the degree of formalization of gender, as represented by obligatoriness of gender agreement and noun classification, as well as by number of agreement targets. The degree of formalization in gender and classifier languages has been investigated by Passer (2016b). Passer compiles two indexes consisting of seven features each, measuring the “Dimen-

sion of Form” and the “Dimension of Transparency” of gender and classifier systems. These indexes are used to investigate the degree of grammaticalization of systems of nominal classification (classifiers and gender; for gender, which he defines very narrowly, he uses the term “concord”). Passer argues that conventionalization (reducing transparency) and formalization can be conceived of as independent pathways of systems of nominal classification. With its 37 systems from 36 languages, Passer’s sample is not particularly large, but it has the advantage that it has world-wide scope, is stratified and also comprises both gender and classifier systems. Passer takes for granted that the Form features and the Transparency features form two dimensions, but the extent to which the features cluster can actually be tested on the basis of Passer’s database. Figure 4 shows a hierarchical clustering of a comparison of the ranking of the 14 features and the two indexes with squared Spearman’s Rho (which is equally sensitive to positive and negative correlations; `varclus()` in the R `Hmisc` library described in Harrell 2001).



Legend: x total formalization value, x1 inventory size, x2 host number (targets within NP), x3 locus operandi (targets outside of the NP), x4 obligatoriness, x5 boundedness, x6 multiple marking on various types of targets in the NP, x7 exhaustivity of classification, y total transparency value, y1 degree of semantic assignment, y2 number of different assignment rule types, y3 number of assignment rules, y4 INDEPENDENCE FROM OTHER GRAMMATICAL CATEGORIES, y5 discreteness of markers, y6 redundancy, y7 flexibility

Figure 4: Clustering of Passer’s (2016b) Form (x) and Transparency (y) features

Figure 4 suggests that there are actually more than two dimensions and that the total indexes do not reflect all of their components equally well. The three first transparency features (y1–3), and apparently also the whole y-index, mea-

sure similar things, viz. how transparent assignment is, ranging from semantic to opaque. Degree of formalization (x) seems indeed to be an important issue, but, as it turns out, y4 (in)dependence of other grammatical categories – even though arguably indicating transparency – actually correlates with multiple marking on various types of targets in the NP (x6), obligatoriness (x4), and boundedness (x5), which seems to indicate that it is a characteristic property of a grammaticalized category of gender to exhibit interdependence with other grammatical categories.

We do not want to suggest here that gender does not exist if it does not cumulate with number, case, or person. However, where there is gender and no cumulation, gender tends to have a low degree of formalization. Notably, gender has a tendency not to be obligatory and not to be marked on multiple agreement targets, if it does not cumulate with other categories. Let us consider a few cases in point.

Within Sino-Tibetan, Limbu (van Driem 1987: 21) and other Kiranti languages (Ebert 2003b: 508) have a very limited masculine-feminine gender opposition on attributive adjectives (one target type). The suffixes, masculine *-pa/ba* and feminine *-ma*, although they can be shown to be of nominal origin (*ma* and *pa* also mean ‘mother’ and ‘father’, for instance, in Camling; Ebert 2003a: 535), are common derivational suffixes in adjectives throughout Tibeto-Burman languages. In Classical Tibetan (Sino-Tibetan, Bodish), adjectives have nominal suffixes (*-pa/-ma/-po/-mo* or *-ka*): *chen-po* ‘large’, *legs-pa* ‘good’, *gsha-ma* ‘worthy’; “a few adjectives may express the natural gender of their referent by alternating the masculine *pa/po* and feminine *ma/mo* suffixes, but most adjective forms are fixed” (DeLancey 2003: 373). Not all adjectives, where the markers occur, do agree and agreement is not obligatory even in those adjectives where it occurs.

The Hindu Kush Indo-Aryan languages Khowar, Kalasha, and Dameli (Liljegrén 2019 [in Volume I]) distinguish animacy in the root of the copula. Number and person are marked through suffixes attached to the animate/inanimate roots, and thus do not cumulate with the morpheme where animacy is marked.

Mopán Maya masculine and feminine gender (originating from person name markers) have only a single marking target. Only a minority of nouns are gendered and the gender marker can sometimes be omitted (Contini-Morava & Danziger 2018: 133).

In Ngan’gityemerri (discussed in §4.3, included in Passer’s sample), gender does not cumulate and not all nouns are classified in noun classes.

In languages where gender has been borrowed, gender is often not in cumulation with another grammatical category and not obligatory. For instance,

Chamorro has borrowed the Spanish masculine and feminine gender markers as *-o/u* and *-a* along with Spanish words which results in a semi-productive sex-based type of gender system without cumulative exponence (Stolz 2012; Di Garbo & Miestamo 2019 [this volume]).

However, before hastening to conclusions, it is important to note that degree of formalization has played an important role in delimiting gender from classifiers. Notably, obligatoriness is a traditional feature used for distinguishing between classifiers and gender (e.g., Dixon 1982: 160: “a grouping of all the nouns of a language [...] so that there is some overt indication of the class of a noun within any sentence in which it occurs”). According to these criteria, most of the languages discussed in this section would count as lacking gender. By applying these criteria, there is thus a danger of excluding by definition languages where gender has limited degree of formalization (see also Wälchli 2019 [this volume]). Yet, we have chosen to make the connection between gender and cumulation explicit in our definition of gender, which contains the statement that gender typically exhibits cumulative exponence with number, case, and/or person (see §1). However, this does not mean that categories lacking cumulation with other categories should be excluded from the study of grammatical gender.

8.4 Summary

To sum up, cumulation of gender with number, case and/or person is pervasive across the languages of the world. In addition, in a few cases we are able to establish through diachronic comparison that cumulative exponence with other morphosyntactic features can be reconstructed, and thus exists from the very origin of the history of a language- and/or family-specific gender system. This can most likely be explained with the fact that gender tends to develop from pre-existing grammatical systems. For instance, gender may arise as a condition on the distribution of a specific number value (as in animacy-constrained plural marking) or case distinction (as in differential argument marking). More research is needed to explore the diachronic relationship between gender and number, case, and/or person, but it is fair to say that interdependence of gender with these other grammatical category types is the rule rather than the exception. This typological finding is in need of diachronic explanation in each individual instance.

Cumulative exponence is a violation of the Principle of One-Meaning–One-Form (one and the same affix is associated with two or more grammatical meanings) and the Principle of Independence (the encoding of gender distinctions is dependent on number, case, and/or person values) and thus qualifies as a phe-

nomenon that fosters complexity (see §2). The fact that gender typically has cumulation with other nominal morphosyntactic features naturally means that gender is usually complex.

9 Lexical plurality and grammatical gender

9.1 Introduction

In many languages, certain nouns tend to be inherently specified for number, that is, to display lexicalized number values. This fact has been claimed to blur the boundaries between the gender and number domain. The gender systems of the Papuan languages Coastal Marind and Walman, described by Olsson (2019 [in Volume I]) and Dryer (2019 [in Volume I]) are a case in point, which we discuss in this section in the light of the larger typological context.

The label *pluralia tantum* is typically used in the literature to refer to nouns that only exist in the plural-marked form, as in English *scissors*, *trousers*, *leftovers*, and *supplies*.²⁶ Broadly speaking, *pluralia tantum* nouns fall within the wider domain of LEXICAL PLURALITY. The term encompasses a variety of semantic and formal phenomena, both morphological and syntactic, which stem from the fact that plurality is a lexicalized property of a given noun, or, simply put, part of what there is to know about it (Acquaviva 2008: 2). In this section we use the labels LEXICAL PLURALS and LEXICAL PLURAL NOUNS as general terms to refer both to *pluralia tantum* nouns, that is, nouns with fixed plural number, as well as to nouns that are inherently plural, but that are not necessarily marked as plural.

Previous studies both on the spoken and signed modality (Koptjevskaja-Tamm & Wälchli 2001; van der Meer 2015; Börstell et al. 2017) show that some broader semantic domains may be identified as recurrent attractors of lexical plurality across languages, while languages differ considerably with respect to the specific concepts that tend to be associated with lexical plurality. In Table 17 we list some major semantic domains – they need not necessarily exclude each other – that have been shown to be most typically associated with lexical plurality. We illustrate each semantic domain with one exemplar concept with an English label. Notice that the concept chosen to exemplify a particular semantic domain need not to be a lexical plural of English, which is indicated by small caps.

Concepts typically expressed by lexical plurals differ in whether they are countable or non-countable, a distinction that not necessarily neatly aligns with

²⁶For a recent, typologically informed, classification of types of *pluralia tantum* nouns, see Corbett (2018).

Table 17: Semantic domains associated with nominal lexical plurality across languages

Semantic domain	Exemplar concept
Abstract ^a	ANGER
Collectives	CATTLE
Dual entities/Internally complex concept	GLASSES
Disease	MEASLES
Festivities and time intervals	SEASON
Liquids and masses	SALIVA
Locations	WOODS
Situations/activities involving more than one participant	FIGHT

^aAbstract nouns can be either count or non-count and it is reasonable to suspect that it is the latter type that is especially likely to be attracted by the lexical plurality domain. We are grateful to Östen Dahl for this suggestion.

the domains in Table 17. Countable units may refer to what we may think of as singular entities. The English nouns *leftovers* and *supplies* have a mass noun reading and it is not possible to talk about one item of them. Liquids and masses are usually non-countable, but also abstract concepts often belong to this category. Conversely, we can talk about *a pair of scissors/trousers*, which, in this respect, behave as count nouns. Languages differ as to whether they use special constructions to count multiple instances of a particular entity denoted by countable lexical plural nouns (as in English *one pair of scissors/trousers*), a topic which is not further addressed here.

As mentioned above, in spoken languages, nouns that are lexically plural typically only occur in the plural form. A parallel situation is found in the signed modality where lexical plurality is associated with double-handed signs, what Börstell et al. (2017) refer to as **ARTICULATORY PLURALITY**. Similarly to the spoken modality, where pluralia tantum nouns are typically marked by regular, productive number morphology, double-handed articulation is used in sign languages to mark non-lexical, compositional plurality with various types of signs (Carl Börstell, p.c.).

9.2 Lexical plurality and grammatical gender: a crosslinguistic overview

If a language has number agreement, lexical plural nouns typically trigger plural agreement, and the formal marking patterns are typically indistinguishable from those triggered by morphological plurals. While this would seem to be a rather unproblematic fact, it turns out that in languages with grammatical gender and large classes of lexical plural nouns, lexical plurality may come to interact so closely with the morphosyntactic encoding of gender that the two domains (gender and number) may appear to be merged into one. This is the situation that we encounter in two of the Papuan languages investigated in the first volume of this work, Coastal Marind (Olsson 2019) and Walman (Dryer 2019). Let us briefly summarize the Coastal Marind and Walman situations (for more extensive analyses, we refer to the individual chapters).

There are four genders in Coastal Marind: masculine, feminine, and two inanimate genders, which Olsson refers to as gender I, II, III, and IV. While gender I and II vary according to number (singular and plural), the two inanimate genders are number-invariant. In addition, the plural marker used for the two animate genders (I and II) and the marker of gender IV are the same, and this syncretism is systematic across all agreement targets, even through the patterns of suppletion that regulate argument indexing on verbs. While male humans are gender I and female humans gender II, there are no strong tendencies that help predict which inanimate nouns should be assigned to gender III and which other ones to gender IV. Nevertheless, some regularities can be detected. For instance, some of the semantic domains that are typically associated with lexical plurality tend to cluster in gender IV (e.g., internally complex objects, diseases, heterogeneous objects). This, together with the fact that the plural marker of the animate genders is systematically syncretic with the marker of gender IV, suggests that there might be an even tighter relationship between gender IV and nominal plurality. According to Olsson, this relationship can be understood in diachronic terms. He speculates that at least some of the gender IV nouns are the diachronic descendants of a large class of pluralia tantum nouns which, as such, used to trigger semantically motivated plural agreement, the same plural agreement pattern triggered by animate masculine and feminine nouns. Such an originally coherent class of pluralia tantum nouns later expanded “resulting in a large, semantically heterogeneous residue gender, with a small core that still reflects the ‘plural semantics’ of the original pluralia tantum grouping” (Olsson 2019 [in Volume I], p. 219).

Walman has two clear-cut gender values, masculine and feminine. In addition, together with the diminutive, Dryer (2019 [in Volume I]) describes lexical plural nouns as a gender-like phenomenon. Lexical plurals in Walman are not marked as plural, but can be described as syntactically pluralia tantum nouns because, independently of whether their denotational meaning is singular or plural, they always trigger plural agreement. Semantically, the range of meanings expressed by lexical plural nouns in Walman strongly overlaps with the semantic groupings identified in the typological literature on the topic: objects consisting of multiple parts, dual entities (especially body parts coming in pairs), mass nouns. While there can be mismatches (not all mass nouns are, for instance, pluralia tantum), the semantic makeup of this class of nouns is highly consistent. According to Dryer, what could justify describing the Walman pluralia tantum nouns as an independent gender value, alongside masculine and feminine, is the sheer number of lexemes in this class: 81 instances of lexical plurals are attested in Dryer's corpus, as opposed to 40 instances of masculine nouns.

Interactions between lexical plurality and grammatical gender similar to those attested in Coastal Marind and Walman are also found in other New Guinean languages. An interesting parallel to Coastal Marind is, for instance, the Ok language Mian. In Mian, along with the masculine and feminine genders there are two inanimate genders: neuter 1, which is sensitive to number distinctions, and neuter 2, which is number-invariant and whose marker is the same as the plural of neuter 1 (for a detailed description of the gender system of Mian, see Fedden (2011). Olsson (2019 [in Volume I]) notes that in Mian the overlap between neuter 2 nouns and the semantic domains typically associated with lexical plurality is even stronger than in Coastal Marind.²⁷ In his survey of gender systems in the languages of New Guinea, Svärd (2019 [in Volume I]) mentions the case of another New Guinean language, Ama (Left May), where lexical plurals systematically align with one gender value in a way that is at least partially reminiscent of the Coastal Marind system. There are three genders in Ama – masculine, feminine, compound – and nouns that are semantically connected with lexical plurality (in particular, nouns denoting objects having many parts and mass nouns) are always assigned to the compound gender (Årsjö 1999: 68). In sum, the language-specific and cross-linguistic data presented in separate contributions to the two volumes of this work show that a number of genealogically unrelated New-Guinean languages have classes of nouns, which fall in between

²⁷Depending on the genealogical classification adopted, Anim and Ok, the language families to which Coastal Marind and Mian, respectively, belong, may be also seen as distantly related members of the Trans New Guinea phylum (see §11.1).

representing a proper gender value and an unusually large class of nouns with fixed plural number/lexicalized plurality. The spread of this pattern within New Guinea and its role as a possible characteristic feature of the gender systems of this area would deserve to be further investigated.

There are a few typological parallels to the New Guinean languages discussed above, where, other things being equal, lexical plurality has an impact on patterns of encoding in the domain of gender and number agreement. One such parallel is Cushitic languages, or at least a subset of them. Cushitic languages are a branch of the Afro-Asiatic family, spreading from Eritrea all the way down to Tanzania and consisting of approximately 40 languages, further divided into four subgroups: Agaw, Beja, East Cushitic, and South Cushitic. Nominal number marking in Cushitic is typically not obligatory. Speakers can leave nouns unmarked for number or use a variety of derivational suffixes and/or morphophonological strategies to mark a noun as singular or plural. In the literature on Cushitic languages, number-unmarked nouns are referred to as nouns with general number or as transnumeral nouns, whereas the derivational singular and plural morphemes are labeled as singulative and plurative. Cushitic languages typically have sex-based gender systems with a masculine-feminine distinction. Yet, some languages of the family are described as having three genders, with the third gender class beyond masculine and feminine being traditionally referred to as “the plural”. There are two main scenarios under which some Cushitic languages have been analyzed as displaying a tripartite gender system with a distinction between masculine, feminine and plural gender.

Under the first scenario, languages have agreement patterns that are used to signal that the controller is plural, but that are not used with all plural controllers. This is, for instance, the case of the East Cushitic language Baiso. The gender and number agreement system of Baiso has been described in detail by Corbett & Hayward (1987) and Corbett (2000). In the following, we provide a brief overview of its most relevant properties.

Baiso has two gender distinctions in the singular, masculine and feminine. Verbs agree in gender and number with the subject. With the majority of plural-marked nouns, the agreement pattern triggered by the verb is the same as the one triggered by masculine singular nouns, irrespective of whether the noun is masculine or feminine. This is illustrated in examples (75) and (76), which show gender and number agreement with masculine singular and plural nouns, and feminine singular and plural nouns, respectively.

- (75) Baiso (Afro-Asiatic, East Cushitic; Corbett 2000: 181): gender and number agreement with masculine nouns

a. *lúban* *hudure*
lion(M).GENERAL slept.M
‘The lion slept.’

b. *luban-jool* *hudure*
lion-PL slept.M
‘The lions slept.’

- (76) Baiso (Afro-Asiatic, East Cushitic; Corbett 2000: 182): gender and number agreement with feminine nouns

a. *kimbír* *hudurte*
bird(F).GENERAL slept.F
‘The bird slept.’

b. *kimbir-jool* *hudure*
bird-PL slept.M
‘The birds slept.’

In addition to the two verb forms exemplified in (75) and (76), Baiso has a third verb form, which is only used when the subject (i.e., the controller noun) is the third person plural pronoun, a noun marked by the paucal suffix or one of the underived nouns listed in Table 18. Because it is used with third person plural pronouns, this third verb form is traditionally glossed as PL, “plural”. The use of the plural verb form with two paucal-marked nouns (one masculine and one feminine) is illustrated in (77).

- (77) Baiso (Afro-Asiatic, East Cushitic; Corbett 2000: 181–182): plural agreement

a. *luban-jaa* *hudureene*
lion-PAUC slept.PL
‘A few lions slept.’

b. *kimbir-jaa* *hudureene*
bird-PAUC slept.PL
‘A few birds slept.’

Table 18: Underived nouns selecting plural agreement in Baiso, adapted from Corbett & Hayward (1987: 9)

Semantic groupings	Nouns
Body parts	<i>ilkoo</i> ‘tooth, teeth’; <i>kalaljaa</i> ‘kidneys’; <i>luḳḳaa</i> ‘foot, feet, leg(s)’; <i>iḷ o</i> ‘eye(s)’; <i>ogorroo</i> ‘hair’; <i>moo</i> ‘hips, lumber region’
Collectives	<i>saé</i> ‘cattle’
Mass nouns	<i>eenoo</i> ‘milk’; <i>soo</i> ‘meat’; <i>udú</i> ‘faeces’
Objects coming in pairs	<i>keferoo</i> ‘sandals’

The eleven nouns in Table 18 always select the plural verb form. The noun for ‘kidneys’, *kalaljaa*, can be described as a “paucal tantum” noun as it is only attested in the paucal-marked form (Corbett & Hayward 1987: 9). The suffix *-oo*, in which many of the nouns listed in the table end, is a productive plural suffix in several Omo-Tana languages, a subgroup within East Cushitic to which Baiso also belongs. However, *-oo* is not a productive plural suffix in Baiso (Corbett & Hayward 1987: 19). Within Cushitic studies, the agreement pattern illustrated in (77) has been analyzed as the morphosyntactic realization of a third gender, the plural gender (Mous 2008: 146). The analysis is motivated by the fact that the nouns listed in Table 18 select plural agreement even though they are morphologically underived for number. For these nouns, plurality is a lexically specified feature as masculine and feminine are for other nouns. Corbett & Hayward (1987) and Corbett (2012) describe the peculiar agreement preferences of the nouns listed in Table 18 as lexical exceptions and reject the analysis of plural as a gender value. Semantically these nouns tend to denote collectives (‘cattle’), entities that are prone to occur as pairs (‘kidneys’), or masses (‘meat’). They always select plural agreement because they are semantically and lexically plural. Di Garbo (2014: 121–127) develops this line of reasoning one step further and describes Baiso as a language with a split system of number agreement. While the majority of nouns that undergo regular morphological plural marking do not trigger dedicated plural agreement but an agreement pattern that is syncretic with the one triggered by masculine singular nouns, as in examples (75) and (76), dedicated plural agreement is used only with a closed set of controllers: plural pronouns, paucal-marked nouns, the lexical plurals and a handful of plural-marked nouns that tend to denote small groups. Di Garbo (2014) speculates that the split number agreement system attested in Baiso is semantically motivated and that

the controllers of dedicated plural agreement rank higher on a scale of semantic plurality than derived plural nouns.

There is yet another profile of languages within Cushitic that has been analyzed as displaying a tripartite gender system with plural as a gender value along with masculine and feminine. These are languages that have dedicated patterns of plural agreement that are used with all plural controllers: third person plural pronouns, derived plurals (that is, nouns that are morphologically marked as plural), and nouns that are unmarked for number but nevertheless control plural agreement. In these languages, gender distinctions are always neutralized in the plural. In addition, in these languages nouns that are number-unmarked but that always trigger plural agreement constitute a rather large lexical class. This large class of inherently plural nouns encompasses both typically lexical plural concepts and concepts that are not associated with lexical plurality, somewhat similarly to nouns of gender IV in Coastal Marind. An example of such a language is Konso, an East Cushitic language spoken in Ethiopia. Konso displays subject agreement on the verb, which has three different inflectional forms depending on whether the subject argument is masculine (78a), feminine (78b) or plural (78c and d). The masculine and feminine forms are used if the subject is singular, the plural form is used if the subject is a plural-marked noun (78c) or a noun that is lexically specified as plural (78d). Definite markers, which are suffixed to nouns, only distinguish between singular and plural. The plural form of the definite marker is used both with overtly plural-marked nouns and with nouns that are lexically specified as plural.

(78) Konso (Afro-Asiatic, East Cushitic; adapted from Tsegaye 2017: 36–37):
gender and number agreement

- a. *ɕmayta-siʔ i=kutiʔ-ay*
old.man-DEF.SG 3=sit.down-PFV.3M
‘The old man sat down.’
- b. *aleeta-siʔ i=piʔ-t-i*
hut-DEF.SG 3=fall-3F-PFV
‘The hut fell.’
- c. *laha-dɔʔ-siniʔ i=muk-i-n*
ram-PL-DEF.PL 3=sell-pass-IPFV.FUT-3PL
‘The rams will be sold.’

- d. *flaa-sini?* *i=pat-i-n*
 comb-DEF.PL 3=be.broken-PFV-PL
 ‘The comb disappeared.’

Orkaydo (2013) and Tsegaye (2017) analyze the plural agreement pattern, as realized on verbs and definite markers, as the morphosyntactic manifestation of a gender value. According to this analysis, plural-marked nouns are also considered to be plural in gender. The main arguments that Orkaydo (2013) and Tsegaye (2017) present in support of the plural as a gender-value analysis in Konso are: (i) the large number of nouns that are underived for number and only trigger plural agreement²⁸ and (ii) the fact that not all of these nouns are semantically analyzable as instances of lexical plurals.²⁹

The possibility of positing an independent gender value for lexical plural/pluralia tantum nouns has also been defended for Russian by Zaliznjak (1977). Russian has a tripartite gender system with a masculine-feminine-neuter distinction that is further subject to a number of animacy-based conditions. Gender distinctions are neutralized in the plural. Pluralia tantum nouns always trigger plural agreement, irrespectively of whether they refer to singular or plural entities. This is illustrated in (79).

- (79) Russian (Indo-European, Slavic; Corbett 2012: 237)
odn-i *san-i*
 one-PL.NOM sledge-PL.NOM
 ‘one sledge’

In virtue of the properties illustrated in (79), according to Zaliznjak, pluralia tantum nouns in Russian are better analyzed as representing one independent agreement class, and thus one independent gender value.

Corbett (2012: 237–238) notices that *plural-as-a-gender-value* analyses have only been proposed for languages where gender distinctions are systematically neutralized in the plural. This is the case for Russian and indeed this is also the case for Coastal Marind, Walman, Baiso and Konso. In languages where gender distinctions are maintained in the plural, lexical plurals are usually distributed

²⁸Orkaydo (2013: 318–330) lists 471 Konso nouns. Out of these, 92 are classified as being inherently plural (or, following his analysis, plural in gender), 134 as feminine, and 245 as masculine.

²⁹By inspecting the meanings of the 92 nouns classified by Orkaydo (2013: 318–330) as inherently plural we found that more than half of them (about 50) have denotational meanings that align with the most typical semantic domains of lexical plurality (e.g., mass nouns, body parts coming in pairs, names of activities requiring multiple participants).

across different gender values, but still share the properties of carrying only plural morphology and/or of only triggering plural agreement. This is for instance the case of Italian, where the plurale tantum noun *pantaloni* ‘trousers’ is masculine and selects only masculine plural agreement while the plurale tantum *forbici* ‘scissors’ is feminine and selects only feminine plural agreement as in *i pantaloni* ‘the.F.PL trousers’ and *le forbici* ‘the.M.PL scissors’. Analyzing Italian pluralia tantum nouns as belonging to separate gender values would then mean positing at least two different lexical plural genders in the language, one formally overlapping with the masculine plural and one with the feminine plural. Corbett (2012: 237–238) uses this argument to reject the cross-linguistic validity of *plural-as-a-gender-value* analyses. Conversely, he stresses that in languages where gender distinctions are neutralized in the plural, lexical plural nouns are de facto outside the system of gender distinctions because this system is only active in the context of singular number, which they are devoid of. The exceptional agreement preferences of these nouns are thus to be analyzed as a consequence of them being irregular from the point of view of number and not of gender.

While we agree that having or not having gender distinctions in the plural is a relevant typological parameter to take into account when assessing the type of encodings that lexical plurality may trigger in the domain of gender and number agreement, we believe that language-specific analyses where lexically plural nouns are described as making up a gender value of their own *should not* be a priori considered to be fallacious. The descriptive adequacy of language-specific categories should always be distinguished from what is generalizable across languages with the support of typologically adequate comparative concepts (Haspelmath 2010). Arguing, and demonstrating, that lexical plural nouns in some gendered languages exhibit gender-like properties does not amount to say that the lexical plural nouns of all languages with gender should be analyzed as instances of an independent gender value. In languages like Coastal Marind and Konso, the lexicalization of the plural number value and the presence of large classes of nouns with fixed plural number, which only trigger plural agreement, clearly blurs the distinction between the gender and number domain.

9.3 Extreme lexicalization of number values in Kiowa-Tanoan

In addition to the cases mentioned in §9.2, the gender system of yet other languages may be described as being entirely based on the lexicalization of number values.³⁰

³⁰This subsection was written by Bruno Olsson. We are very thankful to Bruno for his general contribution to our discussion of gender and lexical plurality.

The most extreme cases of lexicalization of number values are arguably found in the languages of the Kiowa-Tanoan family of North America, illustrated here with Kiowa data from Sutton (2014: 310) and Watkins & McKenzie (1984: 78). Kiowa distinguishes singular, dual and plural numbers through a combination of suffixation on nouns and indexing prefixes on verbs. Nouns occur in two forms: the unmarked basic form and the inverse form, derived by suffixation. For every noun in the language it must be specified whether the noun occurs in the basic or the inverse form when reference is made to one, two or three or more entities (the labels basic and inverse are specific to the Kiowa-Tanoan descriptive tradition and should not be confused with similar labels in other grammatical traditions). For example, *tógúl* ‘young man’ is used in the basic form for reference to one or two young men, whereas the inverse form *tógú·dó* must be used for reference to three or more young men. This contrasts with the noun *’onsó* ‘feet’, which occurs in its basic form when reference is made to two or more feet, but in the inverse form *’onsôy* when reference is made to a single foot. For other nouns the basic form refers to two instances of the referent, as with *’álv* ‘(pair of) apples’, whose inverse form *’álv·bv* is used to refer to one apple or three or more apples. A fourth type of nouns lacks the inverse form and occurs in the basic form regardless of the cardinality of the referents.

Each noun in the language exhibits the basic-inverse alternation according to one of these four patterns. In the Kiowa-Tanoan literature the four patterns are referred to as noun classes and numbered I-IV (following Wonderly et al. 1954). Nouns in the four superclasses are further divided into subclasses according to their combinatorics with verb prefixes indexing person/number of core arguments. The intransitive third person paradigm consists of four prefixes: singular \emptyset -, dual *ê*-, plural *gyà*- and inverse *è*-. The inverse verb prefix occurs whenever the inverse form of the noun is used, and the singular and dual disambiguate the number reference of nouns in their basic form. It is the behavior of the plural prefix that reveals the need for subclasses. For example, some class II nouns (‘bucket’, ‘saw’, ‘arrow’) trigger the plural prefix when reference is made to three or more entities, while other class II nouns (‘bed sheet’, ‘peg, stake’, ‘peyote, cactus’) trigger the singular prefix when reference is made to three or more entities; these two patterns form subclasses IIa and IIb. When the full range of indexing patterns is taken into account, the total number of subclasses is between 7 (e.g. Watkins & McKenzie 1984) and 9 (Harbour 2008; the difference in granularity depends on whether some marginal patterns are counted as their own subclasses or not).

It is clear from Wonderly et al.’s (1954) use of the term *noun classes* that researchers realized early on that the Kiowa-Tanoan system of number marking

amounts to a form of noun classification. Nichols' (1992: 141) conclusion that "noun classification appears to have arisen out of number agreement in the Kiowa-Tanoan family" explicitly couches this in diachronic terms (an interpretation repeated by Aikhenvald 2000: 377 and Luraghi 2011: 451).

The parallel with languages such as Coastal Marind, Walman, Konso and Baiso is most evident in the class of Kiowa nouns that trigger invariant plural prefixation on the verb regardless of the cardinality of the referent (class IVc in Watkins & McKenzie 1984). According to Harbour (2008: 46) this class consists of objects composed of several parts ('trousers', 'book', 'necklace', 'tepee', 'headdress'; the multi-part semantics are also noted by Merrifield 1959: 270, "a single item is looked upon as having several constituent parts"), granular mass nouns ('flour', 'salt', 'sand') and abstracts ('problem', 'dance', 'word, language'), which echoes the pluralia tantum-like semantics of the nouns discussed for Coastal Marind, Walman, Konso and Baiso. The important difference is that Kiowa takes the lexicalization much further, and requires that every noun in the language be specified for its "inherent number". For some of the Kiowa noun classes this can be expressed straightforwardly as an inherent number value, so that Kiowa *k^hʔdé* 'trousers' (class IVc) is inherently plural, and *ʔálb* '(pair of) apples' (class III) is inherently dual. For other classes the pattern is more complicated, as with *tól* 'peg, stake' (class IIb) which triggers singular verb prefix when the cardinality of the referent is 1, the dual prefix with cardinality 2, but the singular also when cardinality is 3 and higher.

We think that the Kiowa-Tanoan systems of "inherent number" must be considered gender according to the Hockettian conception of gender as "classes of nouns reflected in the behavior of associated words". This also seems to be the contention of Harbour, who – working in the Chomskyan tradition – equates the Kiowa noun classes with Indo-European gender, with the main difference residing in their semantic basis: the former is based on number and the latter on sex. For our purposes, the important point is that Kiowa-Tanoan languages represent the extreme end of a spectrum in which the organization of nominal number in a language can be more or less gender-like. Further towards the other end of the spectrum we find languages such as Coastal Marind, Walman, Konso and Baiso, in which lexicalized number (in this case, plurality) appears to have blurred the line between gender and number to a much lesser degree.

9.4 Summary

We believe that a particularly promising direction of research on the interaction between gender and lexical plurality lies in diachrony and, in particular, in ex-

amining how the encoding of lexical plurality affects the evolution of gender and number agreement systems. Olsson (2019 [in Volume I]) suggests that a plausible explanation for the peculiar configuration of gender IV in Coastal Marind is that this agreement class evolved from a smaller nucleus of pluralia tantum nouns (which selected plural agreement because semantically plural) and only gradually came to include non-plural types of nouns. A similar hypothesis could be tested on Konso and other Cushitic languages exhibiting large classes of lexical plural nouns. Another promising area of investigation in this domain would be taking a closer look at languages like Baiso, where only certain types of agreement controllers, among which the lexical plurals, trigger the use of dedicated plural agreement, whereas the majority of morphologically plural nouns trigger agreement patterns that are syncretic with either masculine or feminine singular agreement. These languages, where, synchronically, there seems to be a split in the agreement patterns associated with nominal plurality, offer an interesting test case for hypotheses about the evolution and grammaticalization of number agreement, a topic that goes beyond the scope of the present volume.

10 System evolution

10.1 Introduction

System is probably the most commonly unexplained term in the literature on grammatical gender and thus arguably rather void of meaning. However, in this section, we will argue that the notion of system is highly important from a developmental point of view. Furthermore, the relationship between complexity and system needs to be sorted out. The Latin adjective *complex* ‘weaved together’ and the Ancient Greek noun *sústēma* ‘(what is) standing together’ are very close in their original meanings. It is thus not surprising that complexity is often understood in linguistics as system complexity, which somehow wrongly takes for granted that complexity is necessarily connected to systems, especially if complexity is understood in terms of description length.

A very simple way of defining system in linguistics is to say that it is an opposition of at least two markers, and in this sense gender is always organized in terms of systems. However, this simple definition does not capture many of the systematic properties of mature gender. Gender connects different parts of language structure (one might say that it is always a multiple-interface phenomenon): syntax, semantics, and morphology are always involved. Lexicon is fundamentally involved to the extent that gender is lexical. Even phonology is

sometimes involved, notably if there is phonological gender assignment. Mature gender systems imply a high degree of internal organization and, from a developmental perspective, it is interesting to consider how such complex systems can emerge.

In §10.2 we introduce the notion of co-evolution (a set of more than one diachronic change, which are at least partly dependent on each other), which is crucial for processes of system emergence. In §10.3 we discuss various approaches dealing with contextualization of variability where variation that is not accounted for is remotivated. In §10.4 we will argue that co-evolution in both rise and reduction of gender can take the form of cascades of anomalies.

10.2 Co-evolution

Diachronic processes, such as sound change, analogy, reanalysis, grammaticalization, and semantic shift, are often viewed as individual changes. One sound, morpheme, construction or meaning turns into another sound, morpheme, construction or meaning. However, changes can also co-occur in a sequence of connected events. The probably best-known example are push and drag chains of several sound changes that co-determine each other, such as the great vowel shift in English. Since gender consists of systems of at least two markers, individual diachronic processes are usually not sufficient for the modelling of the emergence and evolution of gender. Of course, it cannot be excluded that several changes that may result in a gender system co-occur accidentally, but more often than not there will be some sort of co-evolution of several changes in the evolution of gender.

Even a maximally simple gender system, such as the Japanese (Japonic) grammatical anaphors, *kanojo* ‘she’ (from the attributive form of the obsolete distal demonstrative in its attributive form *kano* plus the Sino-Japanese form *jo* for ‘woman’) and *kare* ‘he’ (from the independent form of the obsolete distal demonstrative; see Ishiyama 2008 and Wälchli 2019 [this volume]), is difficult to imagine without some sort of co-evolution. It is true that the loss of the distal demonstrative series *kano/kare* is a shared development that is important for rendering both forms opaque, but the forms are still heterogeneous. One is a complex NP, the other one is just a simplex demonstrative form. The development of *kare* to masculine ‘he’ presupposes a semantic shift of narrowing to masculine, and this process is hard to imagine without co-evolution of a parallel feminine form that makes that narrowing possible.

It is thus not surprising that the general literature on grammaticalization, which focuses on individual cases of grammaticalization, says very little about

the origin of gender. Heine & Kuteva (2002) only list a few cases such as MAN ('man', 'male', 'person') > THIRD PERSON PRONOUN in ‖Ani (Khoe-Kwadi, Khoe), Lendu (Central Sudanic, Lenduic), and Zande.

10.3 Contextualization of variability

In a system, there are markers and a division of labor among them. It is a reasonable assumption that the markers (often of rather heterogeneous origin) are there first and that the division of labor is put into place in a second step. Here we will discuss two approaches that can help us understand how this can happen: Lass' (1990) concept of exaptation and the experimental research on iterated artificial language learning by Kirby and Smith and collaborators (Kirby et al. 2008; Smith & Wonnacott 2010).

Lass (1990) borrows the term EXAPTATION from biology where it means the "opportunistic co-optation of a feature whose origin is unrelated or only marginally related to its later use" (Lass 1990: 80), such as when the dinosaur ancestors of birds happen to have feathers which later turn out to be useful for flying. Linguistic exaptation is the development by which junk that is kept (instead of being relegated) is later used for some other purpose. Lass (1990) discusses the following two examples. (i) Indo-European distinguished perfect and aorist in the past, a distinction which was lost in Germanic, where the perfect and aorist stem forms were redeployed as singular and plural past stems in strong verbs. (ii) The Dutch alternation between suffix *-e* and Zero in attributive adjectives expressing gender and number agreement was redeployed in Afrikaans as an expression of various classes of adjectives (among other things, simple versus complex/compound adjectives).

Smith & Wonnacott (2010) use iterated learning modelled in an experiment as a tool for investigating the cultural evolution of language. One group of participants is presented with some stimuli they have to learn and the next group of participants has to learn the language reproduced by the first group and so on in several "generations". The equivalent of Lass' "junk" is free variation in the input. In Smith & Wonnacott's (2010) experiment, learners were presented with nouns denoting animals with the two artificial plural words *fip* and *tay* distributed entirely randomly in the input for the first "generation". This junk, or pattern of free variation between two plural marking strategies, was redeployed in iterative learning. Smith & Wonnacott's (2010) call this PROBABILITY-MATCHING BEHAVIOR: the learners reproduce markers more or less with the same proportion of frequency that the markers have in the input. However, as a consequence of transmission over several generations, the distribution of markers is made

predictable by linking it to particular conditions, in this case the use of markers is made predictable by lexical conditioning. “A typical fifth-participant language exhibits [...] predictable variability [...] for instance, *fip* used to mark plurality on *cow* and *pig*, *tay* used to mark plurality on *rabbit* and *giraffe*” (Smith & Wonnacott 2010: 447). The learners thus developed some sort of lexical gender. While the token frequency of markers changes very little, there is a change from zero predictability to full or almost full predictability. As a consequence, conditional entropy drops, and if entropy is considered a measure of complexity, complexity drops. (Even though system complexity increases as we go from one grammatical distinction, number, to two, number and gender).

Lass’ and Smith & Wonnacott’s examples have in common that there is a co-evolution of many changes. Parallel changes take place in all Germanic strong verbs, all Afrikaans attributive adjectives and all nouns denoting animals in the experiment. Unmotivated alternations are conditioned, which makes the alternation predictable (lower complexity as meaning and form are better aligned) at the cost of a lower independence (higher complexity according to the Principle of Independence), while the number of markers remains constant.

In §7.5 and §8 we have seen that gender may emerge as a condition on an already existing grammatical category. This may seem strange if viewed as a complexification in terms of the Principle of Independence without any obvious benefit since grammatical gender does not seem to provide any communicative benefit. However, rise of gender is better understandable if we assume that the stage before there was gender contained some markers whose use was largely unpredictable. In more general terms, we can assume that the stages that precede the development of gender contain anomalies where some formal distinctions are poorly motivated. This can, for instance, be due to sound change, to decategorialization of nouns, or to anaphoric NPs having become opaque (as in Japanese).

10.4 Reduction and rise of gender as cascades of anomalies

Gender system evolution often involves a sequence of changes where the first change introduces increasing complexity in the form of unpredictable variability and subsequent changes restore order. Such an initial change introducing idiosyncratic patterning can be regular sound change. A well-studied example is the loss of gender agreement in the predicative adjective (but not in the attributive adjective) in German (Fleischer 2007a, Fleischer 2007b and the literature surveyed there). Old High German and Old Saxon had two competing inflectional paradigms of adjectives, one with endings originating from the pronom-

inal paradigm and one with nominal endings. The nominal endings happened to be reduced to zero by regular sound change in all three genders in the nominative singular and in the nominative neuter plural. The idiosyncratic distribution created by phonological erosion is reflected quite accurately in Old Saxon in predicative use (SG 0%, M.PL 99%, F.PL 95%, N.PL 29%; Fleischer 2007a: Table 9). In Early Old High German, two opposite tendencies can be observed in predicative use. On the one hand, inflection tends to be lost in the forms where it was preserved. On the other hand, inflection is also partly reintroduced by analogy to the forms where it was not lost by sound law. Inflected forms spread most easily to the neuter plural and to a lesser extent also to the feminine singular, which happened to have the same pronominal ending as the neuter plural (N.SG 0%, M.SG 1%, F.SG 8%, N.PL 64%, F.PL 79%, M.PL 80%; Fleischer 2007a: Table 11). While the uninflected forms were generalized in predicative use in Middle High German and Modern German, the inflected forms were generalized in Highest Alemannic dialects with support of language contacts with Romance languages (Fleischer 2007b). In attributive use, the inflected pronominal forms with gender and number agreement were generalized in all varieties of German.

In the development simulated by Polinsky & Van Everbroeck (2003) for the transition from Latin to Old French, “the major push for the restructuring of the gender system came from phonological changes (loss of vowel length, loss of word-final segments)” (Polinsky & Van Everbroeck 2003: 385). Neuter merged with masculine in the singular and with feminine in the plural (as preserved in Romanian). In early Old French text, Romanian-like neuter nouns had been reduced to about 4.6% as compared to 21.1% neuter in Classical Latin.

There are also cascades of changes where an anomaly is remedied by restructuring which entails another anomaly which again calls for restructuring which in its turn is an anomaly and so on. Such a cascade of changes is responsible for a strange pattern in some Tamian Latvian dialects in northern Kurzeme where demonstratives do not agree in gender anymore (only in number and case) and always take the masculine form (80).

- (80) Kandava Latvian (Indo-European, Baltic; Graudiņa 1958; Rudzīte 1964: 65; Wälchli 2018: 144)

un tas cūkgans a visām tiēm
 and that.NOM.SG.M swineherd(M).NOM.SG with all.DAT.PL.F that.DAT.PL.M
cūkam tur i palic:s.
 swine(F).DAT.PL there be.PRS.3 stay.PST.PTCP.ACT.NOM.SG.M
 ‘and this swineherd had remained there with all those pigs’

The starting point is a regular sound change (triggered by language contact with the Finnic contact language Livonian) where short vowels in final syllables of words longer than one syllable are lost. This causes gender neutralization (of masculine and feminine) in the accusative plural in nouns. Demonstratives are monosyllables and monosyllables are not affected by the sound change entailing neutralization. However, the neutralization is extended to them by analogy. The masculine accusative plural form in demonstratives is generalized also with feminine controllers. Since there is a syncretism of feminine plural accusative and nominative, the use of masculine forms instead of feminine is extended also to the nominative plural, which causes the gender opposition in the demonstrative plural forms to be maintained only in the dative (attested in the dialect of Zlėkas). This is a new anomaly, the dative is less frequent than the nominative; thus masculine is further expanded to all plural forms in the demonstrative (attested in Puze and Pope). Demonstratives are the only target in these varieties that inflects for gender only in the singular and not in the plural. This is still an anomaly. In the dialect of Dundaga, the generalized use of masculine forms in demonstratives is further extended to all case-number forms of the demonstrative (see Wälchli 2017).

Wälchli (2018) considers the rise of gender in Nalca from the point of view of system emergence. The development in Nalca implies a large number of minor changes of different kinds (grammaticalization, analogy, and reanalysis) that all must have taken place within a short period of time. There are instances of grammaticalization (female person name marker *ge* from *gel* ‘woman’), instances of reanalysis (*nimi ara* [men TOP] > *nim e-ra* [men DN-TOP]), and instances of analogical extension such as when gender is extended to the comitative postposition (*be-b/ge-b/ne-b/e-b/a-b* instead of just *ab* as in other Mek languages). Most of these developments are highly language-specific and are triggered by local anomalies that give rise to new anomalies which again trigger further changes. As a whole, the development in Nalca is a highly specific development, which gives rise to a gender system with highly specific properties. However, since gender systems often exhibit highly specific properties, it can be assumed that complex system emergence of the kind that it can be reconstructed for Nalca may have taken place in other gender systems as well.

11 Areal and genealogical patterns and external factors

In this section, we will discuss patterns in gender that go beyond language-internal implications. §11.1 deals with genealogical and areal patterns. §11.2 addresses external factors in the ecology of languages.

11.1 Areal and genealogical patterns

If we take the nine language families in the world with more than a hundred languages (according to Hammarström et al. 2018), gender can arguably be reconstructed for the proto-language in three of them (Atlantic-Congo, Afro-Asiatic and Indo-European), which testifies to the diachronic stability of gender. However, in all three families there are also a considerable number of languages that have lost gender. And, at least if we adopt a broad definition of gender, the remaining six large language families (Austronesian, Sino-Tibetan, Nuclear Trans-New Guinea, Pama-Nyungan, Otomanguean, Austroasiatic) all have some languages with gender, and in all six families, gender must have emerged more than once. What contributes to the impression that gender is genealogically stable is its entrenchment in specific morphological marking patterns, which makes gender an interesting feature to look at for traditional historical linguistics. As Nichols (2003: 303) puts it, “[f]or genders, with their clear formal exponents, it is very obviously not the abstract typological feature but particular form-function pairings that are transmitted from ancestor to daughter language”.

However, old morphological material does not necessarily guarantee wide distribution across a large language family. A case in point is gender in Classical Tibetan and Kiranti languages discussed in §8.3, where masculine *-pa/po* and feminine *-ma/mo* are common derivational suffixes in adjectives throughout Tibeto-Burman languages, so it cannot be excluded that gender in Sino-Tibetan might be old.

There is probably a bias toward discussing stable gender in historical linguistics more often than instable gender. This is understandable since only morphologically entrenched stable gender is useful for establishing genealogical groupings of languages. There are so far no general surveys of the development of gender across Austronesian, Sino-Tibetan, Nuclear Trans-New Guinea, Otomanguean or Austroasiatic (for Australian languages, however, see Dixon 2002: 449–514), and no general surveys for the loss of gender across Atlantic-Congo, Afro-Asiatic or Indo-European.

Classifiers are more prone to areal diffusion than grammatical gender (see Seifart 2010: 730–731 and the references given there). However, this does not mean that language contact is irrelevant for gender. Nichols (2003: 300) argues that gender is a cluster phenomenon in the sense that it is most easily preserved where languages with gender are neighbors of (usually) related languages with gender. Put differently, gender is “of high stability only when reinforced by gender systems in neighboring languages” (Nichols 2003: 303) and languages that lose gender are typically neighbors of each other. This does not only hold for gender

in general, but also for particular gender agreement targets, as the preservation of gender in predicative adjectives in Highest Alemannic German dialects due to contacts with Romance languages discussed in §10.4 (Fleischer 2007b).

The findings of Liljegren (2019 [in Volume I]) on the distribution of gender in Hindu Kush Indo-Aryan are well in line with Nichols' (2003) suggestion. Liljegren identifies areal patterns both in the loss of gender, but also in the emergence of a new gender opposition based on animacy. Liljegren also highlights the diachronic dimension. The two Chitral group languages, Khowar and Kalasha, which have lost the Indo-Aryan masculine-feminine opposition and developed a new gender system based on animacy are likely to reflect a first wave of Indo-Aryan settlers in the Hindu Kush area. Languages with concurrent sex- and animacy-based systems are spoken in the vicinity of Chitral languages.

Areal patterns in the development of gender within clusters of languages of the same family can also be identified in other areas. Within the Austroasiatic Khasian branch, War-Jaintia is clearly more distantly related to Khasi than Lyngngam based on evidence from lexical data (Nagaraja et al. 2013: 6). However, the similarities of gender systems rather follow areal patterns where the westernmost language Lyngngam (Nagaraja 1996) has the most rudimentary system among the Khasian languages (see also Di Garbo & Miestamo 2019 [this volume]). In Northern Australia, the Ngan'gityemerri Nangikurrunggurr (Southern Daly) nominal classification system is more similar to that of Marithiel (Western Daly) than to that of Murriny Patha, even though Murriny Patha (Southern Daly) is a closer genealogical relative. Marrithiyel and Ngan'gityemerri "share the larger, central classes, have a number of formally cognate classifiers, and display the same range of agreement patterns" (Green 1997: 233). In central New Guinea, Anim and Ok have very similar gender systems (see Olsson 2019 [in Volume I]). They are so similar in form and function that they are likely cognates (Usher & Suter 2015: 118). However, lexical comparison does not suggest any close genealogical relationship of Anim and Ok (E. Suter, p.c.). According to Seifart (2007), the systems of nominal classification in Huitotoan and Boran are so strikingly similar, that entirely independent development is unlikely, but no common proto-system can be reconstructed.

11.2 External factors

As argued by Dahl (2019 [in Volume I]) it is not easily possible to establish any correlations between grammatical gender and ecological parameters, such as population size or degree of contact and there is no positive correlation with morphological complexity (Nichols 2019 [in Volume I]). This contrasts with ev-

idence from other typological features where extralinguistic ecological factors are clearly reflected in typological distributions (Lupyan & Dale 2010; Sinnemäki 2014a). Sinnemäki & Di Garbo (2018) do not find any significant relationship between the number of gender distinctions (including whether or not a language has gender) and sociolinguistic variables, whereas degree of inflectional synthesis in the verb is clearly sensitive to population dynamics. It is, of course, possible that number of genders does not accurately represent the complexity of gender and that other properties of gender systems must be used (for which large scale data sets are not available) to establish a relationship with factors of population dynamics. However, Blasi et al. (2017) do not find any evidence for adaptive patterns in gender marking even when looking at adjectival modifiers and personal pronouns in creole languages. The results from the large-scale quantitative studies conducted so far thus suggest that, if there are correlations between gender typology and sociolinguistic factors, they are rather subtle, so that they are unlikely to be covered in large typological databases.

A problem with large typological databases is that they often do not take into account dialects. The number of genders in Bininj Kun-Wok ranges from four in the central Kunwinjku dialect to zero in Kune, with Gun-djeihmi having three genders. According to Evans (1997), considerable differences in grammatical gender across dialects of Bininj Kun-Wok reflect social relationships with speakers of neighboring languages. In the *WALS* database, the number of genders listed for Bininj Kun-Wok is simply “four”. As Evans (1997: 105) puts it, deep regularities cannot always be seen in the shallow perspective of one dialect. Karatsareas (2014) shows that not all varieties of Koineic Greek are equally conservative, especially not the different varieties of Greek in Asia Minor. In Greek in Asia Minor the number of genders ranges from three (like in Modern Greek in Greece) in Pontic Greek (but with major restructuring of the system) to zero in Cappadocian Greek. Karatsareas (2009) argues that the loss of gender in Pontic Greek results from an interplay of heavy language contact with Turkish and language-internal analogical levellings. Interestingly, dialects of Ancient Greek in Asia Minor not surviving to the present were already undergoing restructuring of their gender systems due to substrate from Anatolian languages, which had only two genders (common and neuter) (Brixhe 1994: 176). As in Greek, in Latvian gender restructuring of very different kinds occur in peripheral dialects with intensive language contact, in this case with Finnic (Livonian and Estonian) (see Wälchli 2017). Like Greek varieties in Asia Minor, the Tamian Latvian dialects are highly endangered.

There is thus evidence from a fair number of particular cases that a large proportion of non-native speakers and/or intensive language contacts with lan-

languages lacking grammatical gender can entail massive restructuring in gender systems which can, but need not, entail a reduction of the number of genders (see also Trudgill 2011: 24). In a study of 36 languages distributed among 15 sets of closely related languages, Di Garbo (forthcoming) finds that in Eurasia radical reduction, loss and emergence of gender agreement tend to cluster around language family edges, which is consistent with the findings of Nichols (2003). “Loss of gender agreement tends to prevail under circumstances in which the demographically dominant and/or more prestigious language lacks grammatical gender. On the other hand, borrowing of gender agreement patterns may be favored when the demographically dominant and/or more prestigious language has grammatical gender” (Di Garbo forthcoming, see also Di Garbo & Miestamo 2019 [this volume]). Prestige of languages with gender also plays a role in cases of language planning as reflected in the gender system of the Makanza variety of Lingala that was designed by missionaries (Meeuwis 2013; see also Di Garbo forthcoming and Di Garbo & Miestamo 2019 [this volume]). Di Garbo (forthcoming) launches the hypothesis that gender marking may actually have important ties to the way in which speakers and speech communities construe their linguistic identity in opposition to that of their neighbors. A case in point is the mixed language Michif which preserves both the gender system of French and the gender system of Cree (Bakker 1997; Di Garbo forthcoming).

12 Conclusions

In this chapter we have addressed grammatical gender and its complexity (as defined in §2) from a dynamic perspective. We found that dynamic comparative concepts of the form *From X to Y*, as summarized in Table 19, are highly useful to describe the typology of gender. Often it is the case that less mature gender is a source for more mature complex gender, which contributes to the view that complexity in gender is something that can grow over time.

Our starting point was a dynamic definition of gender in §1, repeated here for convenience.

Definition of gender adopted in this chapter:

Gender is a grammatical category type with a semantic core of animacy and/or sex reflecting classes of referents, which have a propensity to turn into classes of noun lexemes. It is overtly marked on noun-associated forms. It typically exhibits cumulative exponence with number, case, and/or person. Gender is organized in the form of systems.

Table 19: Less mature gender as source for more mature gender

Simpler earlier stage can develop into a...	...more mature stage	
Referent-based gender >	Lexical gender	§3
Marker in independent use >	Gender in adnominal use	§4
One-to-one assignment >	Many-to-one assignment	§6.2
Semantic gender assignment >	Opaque gender assignment	§6.3
Semantic assignment (“covert” gender) >	Formal assignment (“overt” gender)	§6.4
Morphological assignment or sandhi >	Phonological assignment	§6.4
Classes of single items >	Classes of larger sets	§6.6
Condition on another feature >	Gender feature	§7.5
Apposition and nominal gender targets >	Non-nominal gender targets	§7.5
Non-idiomatic gender >	Idiomatic use of gender	§7.7

This definition goes beyond the traditional Hockettian definition, which is based on two criteria: noun classes and agreement. Our definition is dynamic in the sense that it expresses the fact that gender is an evolving category type, where gender has a semantic core of animacy and/or sex and exhibits hierarchical patterning according to the animacy hierarchy above some cutoff point in the animate segment of the hierarchy (§3). The semantic core and the hierarchical patterning reflect referent-based gender. Gender becomes lexical only as a secondary development. Put differently, the organization of gender in terms of noun classes is a mature phenomenon. Incipient gender need not have noun classes and in the process of gender loss, lexical gender can be lost before referent-based gender is lost. In several language groups, gender can be shown to originate from top segments of the animacy/individuation hierarchy and then move further down the hierarchy as it further develops. Gender thrives in symbiosis with nouns, but does not usually originate as noun classes. When associated with nouns, gender tends to lexicalize. Gender assignment can be semantic, formal, and/or opaque (§5 and §6). Gender has mechanisms to restore semantic assignment for animate referents if gender assignment for animate referents has become opaque (§6.3). In some languages, gender assignment can be flexible. Through flexible gender assignment speakers modify the construal of noun referents, targeting properties such as size and/or countability (§5).

Gender is a special case of nominal marking on noun-associated words where the number of values is larger than one. But there are also many cases of nominal marking with value one without opposition of gender values. Omission of head nouns in NPs and subsequent explicit nominal marking of non-headed NPs

seems to be an important driving force for the accumulation of nominal morphology as the marking of independent modifiers can be transferred to modifiers in headed NPs (§4).

Agreement is complex in the sense that it can involve syntactically complex controllers and syntactically complex targets and in the sense that the relationship between controller and target can be of various kinds: syntactic and strictly intra-sentential, semantic and inter-sentential, or purely contextual in the case of latent controllers. There is always a specific relationship between controller and target in agreement, but this specific relationship need not necessarily be coreference. Features are a highly mature form of agreement and features may develop from conditions. Gender requires displacement (realization on another element than the one triggering it) in order to be considered a grammatical category. Overt marking of gender on nouns is distinct from gender as a grammatical category, and relates to derivation rather than agreement (§7).

Gender systems almost always imply cumulation with number, case and/or person. This is so pervasive that we have decided to include this peculiarity in the definition of gender. Number and case also play an important role for the emergence of gender systems. In general, it seems to be the very essence of gender that it interacts with other grammatical domains, such as number, person, case, and evaluation. To the extent that interaction with other grammatical categories is counted as complexity according to the Principle of Independence, gender is almost always complex. Gender is thus arguably complex by definition. Cumulation with number, case and/or person has not been taken into account sufficiently in the literature pointing out the similarities between gender and classifiers. Classifiers are similar to gender in that they are classes of referents or classes of noun lexemes. However, classifiers do not tend to interact with number and case in the way gender does (§8).

As gender, number can be entrenched in the lexicon in the form of classes of pluralia tantum, and pluralia tantum can further develop into gender values. It is, of course, possible to exclude pluralia tantum from gender by definition, but it is not clear whether this is useful since it is the very essence of gender to be connected to other grammatical categories, and among them number is the most important one (§9).

Gender is organized in terms of systems that connect different parts of language structure (lexicon, syntax, morphology, semantics, phonology) in order to efficiently and orderly assign values to markers. Although the origin of many gender systems is unknown, different kinds of diachronic approaches are indispensable for understanding how gender emerges and evolves as systems (§10).

Gender is stable diachronically in the sense that it is highly entrenched in specific morphosyntactic marking. Gender displays areal patterns especially in groups of closely related languages. Especially in non-mature stages, gender seems to spread across closely related languages or languages with similar typological profiles. Gender is often lost or restructured in languages with intensive contacts with languages lacking gender or displaying different gender systems. There is no obvious general relationship between the typology of gender and language ecology, but larger proportions of non-native speakers and higher population size seem to go together with restructuring in gender marking (§11).

Gender, noun classes and agreement are among the most discussed topics in the linguistic literature, but there are still many open questions which could only be touched upon in this chapter or are not addressed at all. As the literature is growing, there is also a need of integrative surveys, even if only partial ones, like this chapter. We hope that this chapter, and the two volumes as a whole, will stimulate further descriptions of gender in particular languages and dialects, new large-scale typological studies, and more comprehensive surveys of the research than this chapter provides.

Acknowledgments

We would like to thank Östen Dahl, Pernilla Hallonsten Halling, Martin Haspelmath, Don Killian, Bruno Olsson, and three anonymous reviewers for numerous highly useful and thought-provoking comments, which helped us improve this chapter in multiple ways. We are also grateful to Maria Koptjevskaja-Tamm and Björn Wiemer for answering our questions about Slavic languages, and to Sebastian Nordhoff, Felix Kopecky and Martin Haspelmath for their advice and assistance when preparing the final version of the text. Most of all we would like to thank Bruno Olsson, the editor of this chapter, who has done much more for this chapter than you ever might expect from an editor.

Symbols and special abbreviations

The abbreviations and symbols listed below are not found in the Leipzig Glossing Rules:

- { } gender with which the morphological form is more commonly associated
- [] non-overt element
- () inherent category
- I, II, III, ... Genders I, II, III, etc.

A	actor	INT	interrogative (Coastal Marind)
ACT	active		
ACT	actualis (Coastal Marind)	ITER	iterative mood (Meskwaki)
AGN	agentive noun		
AGT	grammatical agent	KA	ka-class (Paumari)
ANIM	animate gender	LNK	linker
AOR	aorist	MAKE	light verb ‘make’ (Oksapmin)
ASS	associative case (Uduk)		
ATTR	attributive (Archi)	N_	non-
CL	class	N1	neuter 1 (Mian)
CL1 etc.	class 1 etc.	N2	neuter 2 (Mian)
CM	common gender; common noun marker (not person name marker: Tagalog)	NARR	narrative
		Neg.indef	negative indefinite pronoun
		NEUT	neutral orientation (Coastal Marind)
Cmpl	complement clause		
COMP	comparative	NF	non-finite (Pnar §4.3)
CV	CV gender (Nalca)	N_M	non-masculine
DERIV	derivation	NOMIN	nominal marker
DETR	detransitivizing “emphatic” form of verb (Bari)	Num	numeral
		N.UNI	non-uniqueness
		OBV	obviative
ECHO	prosodic echo vowel	PAUC	paucal
EMPH	emphatic clitic or particle	PLT	plurale tantum
DIM	diminutive	PN	proper name marker
DN	default noun gender (Nalca)	Poss	possessive pronoun
		POSS	possessive (affix)
DP	default phrase gender (Nalca)	PRED	predicator (Mian)
		PRON	pronominal inflection
GEN	genitive; possession (Pnar)	REAL	realis
		REFL.POSS	reflexive possessive
Gen	noun possessor	Rel	relative clause
GM	gender marker (Mopán Maya)	SPEC	specific
		SUPERL	superlative
HAB	habitual	U	undergoer
HUM	human	UNI	uniqueness
INAN	inanimate gender	WEAK	weak declension
Int	interrogative		

Translations of New Testaments used

- Iraya N. T. 1991. The New Testament in Iraya. <http://listen.bible.is/IRYOMF/Matt/1>.
- Kahua Bible. 2011. Buka Apuna (I Lotu Katolika). The Bible in Kahua. <http://bibles.org/agw-KCB/Gen/int>. Bible Society of the South Pacific.
- Nalca N.T. n.d. Translated by Samuel and Jeremy Souga and Roger Doriot.
- Una N. T. 2007. Translated by Wilem Balyo, Titus Bitibalyo, Melkias Kipka and Gipson Malyo, with consultant input from Dick and Margreet Kroneman (SIL) and Lourens de Vries (UBS). The permission to use the Una N.T. by the Indonesian Bible Society is gratefully acknowledged.

References

- Acquaviva, Paolo. 2008. *Lexical plurals: A morphosyntactic approach*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. 2010. Gender, noun class and language obsolescence: The case of Paumari. In Eithne B. Carlin & Simon van de Kerke (eds.), *Linguistics and archaeology in the Americas: The historization of language and society*, 235–252. Leiden: Brill.
- Aikhenvald, Alexandra Y. 2012. Round women and long men: Shape, size, and the meanings of gender in New Guinea and beyond. *Anthropological Linguistics* 54. 33–86.
- Aikhenvald, Alexandra Y. 2016. *How gender shapes the world*. Oxford: Oxford University Press.
- Aksenov, A. T. 1984. K probleme èkstralingvistièskoj motivacii grammatièskoj kategorii roda [About the problem of the extralinguistic motivation of the category of gender]. *Voprosy jazykoznanija* 1. 14–25.
- Allan, Keith. 1977. Classifiers. *Language* 53(2). 285–311.
- Anderson, Judi Lynn. 1989. *Comaltepec Chinantec syntax* (Studies in Chinantec Languages 3). Arlington: Summer Institute of Linguistics.
- Angerer, Martin. 2009. *Regensburg: Für Anfänger und Fortgeschrittene*. Regensburg: Donau Live.
- Årsjö, Britten. 1999. *Words in Ama*. Uppsala: Uppsala University. (MA thesis).
- Audring, Jenny. 2009. Gender assignment and gender agreement: Evidence from pronominal gender languages. *Morphology* 18(2). 93–116.

- 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.
- 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:10.5281/zenodo.3462756
- Audring, Jenny & Sebastian Fedden. 2018. Introduction. In Sebastian Fedden, Jenny Audring & G. Greville Corbett (eds.), *Non-canonical gender systems*, 3–22. Oxford: Oxford University Press.
- Baker, Mark. 2003. *Lexical categories: Verbs, nouns, and adjectives*. Cambridge University Press: Cambridge.
- Bakker, Peter. 1997. *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis*. Oxford: Oxford University Press.
- Barlow, Michael. 1999. Agreement as a discourse phenomenon. *Folia Linguistica* 33(2). 187–210.
- Becher, Jutta. 2001. *Untersuchungen zum Sprachwandel in Wolof aus diachroner und synchroner Perspektive*. Universität Hamburg. (Doctoral dissertation).
- Beguilot, Francesco. 1942. *Il berbero nefûsi di Fassâto*. Roma: Istituto per l'Oriente.
- Binzell, John. n.d. Nalca original texts. 8 texts collected by John Binzell around 1975. 16 pages, manuscript.
- Blasi, Damian E., Susanne M. Michaelis & Martin Haspelmath. 2017. Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour* 1(10). 723–729. DOI:10.1038/s41562-017-0192-4
- Bond, Oliver, Greville G. Corbett, Marina Chumakina & Dunstan Brown (eds.). 2016. *Archi: Complexities of agreement in cross-theoretical perspective*. Oxford: Oxford University Press.
- Börstell, Carl, Ryan Lopic & Gal Belsitzman. 2017. Articulatory plurality is a property of lexical plurals across sign languages. *Linguisticæ Investigationes* 39. 392–408. Special issue on “Lexical plurals and beyond” edited by Peter Lauwers and Marie Lammert.
- Bosch, Peter. 1988. Representing and accessing focussed referents. *Language and Cognitive Processes* 3(3). 207–232.
- Bresnan, Joan & Jonni Kanerva. 1989. Locative inversion in Chicheŵa: A case study in factorization in grammar. *Linguistic Inquiry* 20(1). 1–50.
- Brixhe, Claude. 1994. Le phrygien. In Françoise Bader (ed.), *Langues indo-européennes*, 165–178. Paris: CNRS Éditions.

- Bybee, Joan. 1985. *Morphology: The study of the relation between meaning and form*. Amsterdam: John Benjamins.
- Capell, Arthur. 1982. *Arosi grammar* (Pacific Linguistics B 20). Canberra: Australian National University.
- Capell, Arthur & Heather H. Hinch. 1970. *Maung grammar: Texts and vocabulary*. The Hague: Mouton.
- Caro Reina, Javier. 2018. The grammaticalization of the terms of address *en* and *na* as onymic markers in Catalan. In Friedhelm Debus, Rita Heuser & Damaris Nübling (eds.), *Linguistik der Familiennamen*, 175–204. Hildesheim: Olms.
- Carstairs, Andrew. 1987. *Allomorphy in inflection*. London: Croom Helm.
- Carter, Hazel. 2002. *An outline of Chitonga grammar*. Lousaka: Bookwords Publishers.
- Chumakina, Marina & Oliver Bond. 2016. Competing controllers and agreement potential. In Oliver Bond, Greville Corbett, Marina Chumakina & Dunstan Brown (eds.), *Archi: Complexities of agreement in cross-theoretical perspective*, 77–117. Oxford: Oxford University Press.
- Comrie, Bernard & Tania Kuteva. 2013. Relativization on subjects. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/122>.
- Contini-Morava, Ellen & Eve Danziger. 2018. Non-canonical gender in Mopan Maya. In Sebastian Fedden, Jenny Audring & Greville Corbett (eds.), *Non-canonical gender systems*, 129–146. Oxford: Oxford University Press.
- Contini-Morava, Ellen & Marcin Kilarski. 2013. Functions of nominal classification. *Language Sciences* 40. 263–299.
- 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. 2012. *Features*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2013. 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. 2018. Pluralia tantum nouns and the theory of features: A typology of nouns with non-canonical number properties. *Morphology* 29(1). 51–108. DOI:10.1007/s11525-018-9336-0

- Corbett, Greville G. & Sebastian Fedden. 2016. Canonical gender. *Journal of Linguistics* 52(3). 495–531.
- Corbett, Greville G., Sebastian Fedden & Raphael Finkel. 2017. Single versus concurrent feature systems: Nominal classification in Mian. *Linguistic Typology* 21(2). 209–260.
- Corbett, Greville G. & Norman Fraser. 1999. Default genders. In Barbara Unterbeck & Matti Rissanen (eds.), *Gender in grammar and cognition*, 55–97. Berlin: Mouton de Gruyter.
- Corbett, Greville G. & Richard J. Hayward. 1987. Gender and number in Bayso. *Lingua* 73(1–2). 195–222.
- Cornish, Frances. 1987. Anaphoric pronouns: Under linguistic control, or signalling particular discourse representations? *Journal of Semantics* 5(3). 233–260.
- Crazzolaro, Joseph Pasquale. 1955. *A study of the Acool language: Grammar and vocabulary*. 2nd edn. London: Oxford University Press.
- Creissels, Denis. 1999. Origine et évolution des diminutifs et augmentatifs dans quelques langues africaines. *Silexicales* 2. 29–35.
- Creissels, Denis, Gerrit J. Dimmendaal, Zygmunt Frajzyngier & Christa König. 2008. Africa as a morphosyntactic area. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 86–150. Cambridge: Cambridge University Press.
- Croft, William. 1994. Semantic universals in classifier systems. *Word* 45(2). 145–171.
- 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. 2005. Word classes, parts of speech, and syntactic argumentation. *Linguistic Typology* 9(3). 431–441.
- 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.
- Dahl, Östen. 2000a. 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. 2000b. Elementary gender distinctions. In Matti Rissanen, Terttu Nevalainen & Mirja Saari (eds.), *Gender in grammar and cognition*. Vol. 2: *El-*

- ementary 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. 2019. Gender: esoteric or exoteric? In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 53–61. Berlin: Language Science Press. DOI:10.5281/zenodo.3462758
- Daladier, Anne. 2011. The group Pnaric-War-Lyngngam and Khasi as a branch of Pnaric. *Journal of the Southeast Asian Linguistics Society* 4(2). 169–206.
- de Hollenbach, Elena E. 1995. Cuatro morfemas funcionales en las lenguas mixtecas. In *Vitalidad e influencia de las lenguas indígenas en Latinoamérica: II coloquio Mauricio Swadesh*, 284–293. Universidad Nacional Autónoma de México.
- DeLancey, Scott. 2003. Classical Tibetan. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages* (Routledge Language Family Series), 253–269. London & New York: Routledge.
- Delbrück, Berthold. 1883. *Vergleichende Syntax der indogermanischen Sprachen*. Erster Theil. Strassburg: Trübner.
- Di Garbo, Francesca. 2014. *Gender and its interaction with number and evaluative morphology: An intra- and intergenealogical typological survey of Africa*. Stockholm University. (Doctoral dissertation).
- Di Garbo, Francesca. 2016. Exploring grammatical complexity crosslinguistically: The case of gender. *Linguistic Discovery* 14(1). 46–85.
- Di Garbo, Francesca. forthcoming. The complexity of grammatical gender and language ecology. In Peter Arkadiev & Francesco Gardani (eds.), *The complexities of morphology*. Oxford: Oxford University Press.
- Di Garbo, Francesca & Yvonne Agbetsoamedo. 2018. Non-canonical gender in African languages: A typological survey of interactions between gender and number, and gender and evaluative morphology. In Sebastian Fedden, Jenny Audring & Greville Corbett (eds.), *Non-canonical gender systems*, 176–210. Oxford: Oxford University Press.
- Di Garbo, Francesca & 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:10.5281/zenodo.3462778

- Di Garbo, Francesca & Annemarie Verkerk. 2018. The evolution of Bantu gender marking systems: Typology, phylogeny and social history. Plenary lecture at the 51th Meeting of the Societas Linguistica Europaea. Tallinn (Estonia).
- Dietze, Joachim. 1973. Die Entwicklung der altrussischen Kategorie der Be-seeltheit im 13. Und 14. Jahrhundert. *Zeitschrift für Slawistik* 18(1). 261–272.
- 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. 1977. *A grammar of Yidjip*. Cambridge: Cambridge University Press.
- Dixon, Robert M. W. 1979. Ergativity. *Language* 55. 59–138.
- Dixon, Robert M. W. 1982. 'Where Have all the adjectives gone?' and other essays in semantics and syntax. Berlin: Mouton de Gruyter.
- Dixon, Robert M. W. 2000. Categories of the noun phrase in Jarawara. *Journal of Linguistics* 36(3). 487–510.
- Dixon, Robert M. W. 2002. *Australian languages: Their nature and development* (Cambridge Language Surveys). Cambridge: Cambridge University Press.
- Donoghue, Emma. 2010. *Room*. New York: Little, Brown & Company.
- 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, accessed 2015-2-6. Unpublished manuscript.
- Dost, Ascander & Vera Gribanova. 2006. Definiteness marking in the Bulgarian. In Donald Baumer, David Montero & Michael Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*, 132–140. Somerville, MA: Cascadilla Proceedings Project.
- Dravniece, Sofija. 2008. *Dundagas izloksnes teksti*. Brigita Bušmane (ed.). Rīga: LU Latviešu valodas institūts.
- Dressler, Wolfgang & Ursula Doleschal. 1990. Gender agreement via derivational morphology. *Acta Linguistica Hungarica* 40. 115–137.
- 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:10.5281/zenodo.3462766

- Ebert, Karen. 2003a. Camling. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 533–545. London: Routledge.
- Ebert, Karen. 2003b. Kiranti languages: An overview. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 505–517. London: Routledge.
- Enger, Hans-Olav. 2004. Scandinavian pancake sentences as semantic agreement. *Nordic Journal of Linguistics* 27(1). 5–34.
- Enger, Hans-Olav & Greville G. Corbett. 2012. Definiteness, gender, and hybrids: Evidence from Norwegian dialects. *Journal of Germanic Linguistics* 24(4). 287–324.
- Epps, Patience. 2008. *A grammar of Hup*. Berlin: Mouton de Gruyter.
- Epps, Patience. 2012. Between headed and headless relative clauses. In Bernard Comrie & Zarina Estrada-Fernández (eds.), *Relative clauses in languages of the Americas: A typological overview*, 191–211. Amsterdam: John Benjamins.
- Erbaugh, Mary. 1986. Taking stock: The development of Chinese noun classifiers historically and in young children. In Colette Grinevald Craig (ed.), *Noun classes and categorization*, 399–436. Amsterdam: John Benjamins.
- Evans, Nicholas. 1994. The problem of body parts and noun class membership in Australian languages. *University of Melbourne Working Papers in Linguistics* 14. 1–8.
- 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. 2000. Iwaidjan: A very un-Australian language family. *Linguistic Typology* 4. 91–142.
- Evans, Nicholas. 2003. *Bininj Gun-Wok: A pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: Pacific Linguistics.
- Evans, Nicholas. 2015. Valency in Nen. In Andrej Malchukov & Bernard Comrie (eds.), *Valency classes in the world's languages*. Vol. 2: *Valency in Nen: Case studies from Austronesia and the Pacific, the Americas, and theoretical outlook*, 1069–1116. Berlin: Mouton de Gruyter.
- Faarlund, Jan Terie. 1977. Embedded clause reduction in Scandinavian gender agreement. *Journal of Linguistics* 13. 153–168.
- Fedden, Sebastian. 2010. Ditransitives in Mian. In Andrej Malchukov, Martin Haspelmath & Bernard Comrie (eds.), *Studies in ditransitive constructions: A comparative handbook*, 456–485. Berlin: Mouton de Gruyter.
- 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
- Fife, James & Gareth King. 1986. Celtic (Indo-European). In Andrew Spencer & Arnold Zwicky (eds.), *Handbook of morphology*, 477–499. Oxford: Blackwell.
- Fischer, Olga, Ans van Kemenade, Willem Koopman & Wim van der Wurff (eds.). 2000. *The syntax of Early English*. Cambridge: Cambridge University Press.
- Fleischer, Jürg. 2007a. Das prädikative Adjektiv und Partizip im Althochdeutschen und Altniederdeutschen. *Sprachwissenschaft* 32. 279–348.
- Fleischer, Jürg. 2007b. Zur Herkunft des flektierten prädikativen Adjektivs im Höchstalemannischen. *Zeitschrift für Dialektologie und Linguistik* 74(2–3). 196–240.
- Foley, William A. & Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Frantz, Donald. 1995. Southern Tiwa argument structure. In Clifford S. Burgess, Katarzyna Dziwirek & Donna Gerds (eds.), *Grammatical relations: Theoretical approaches to empirical questions*, 75–98. Stanford: CSLI.
- Gabelentz, Georg von der. 1891. *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherige Ergebnisse*. Leipzig: Tauchnitz.
- Gerds, Donna B. 2013. The purview effect: Feminine gender on inanimates in Halkomelem Salish. In Chundra Cathcart, I-Hsuan Chen, Greg Finley, Shinae Kang, Clare S. Sandy & Elise Stickles (eds.), *Proceedings of the 37th annual meeting of the Berkeley Linguistics Society*, 417–426. Berkeley: Berkeley Linguistics Society.
- Gore, Edward Cline. 1926. *A Zande grammar*. London: Sheldon Press.
- Graudīņa, Milda. 1958. Laidzes un Kandavas izloksne [The dialects of Laizde and Kandava]. Rīga: University of Latvia Kandidata disertācija [=PhD dissertation].
- Green, Ian. 1997. Nominal classification in Marrithiyel. In Mark Harvey & Nicholas Reid (eds.), *Nominal classification in Aboriginal Australia*, 229–253. Amsterdam: John Benjamins.
- 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.
- Greyerz, Otto von & Ruth Bietenhard. 1997. *Berndeutsches Wörterbuch*. 6th edn. Muri bei Bern: Cosmos.
- Grinevald Craig, Colette. 1977. *The structure of Jacalteco*. Austin: University of Texas Press.

- Grinevald, Colette & Frank Seifart. 2004. Noun classes in African and Amazonian languages: Towards a comparison. *Linguistic Typology* 8(2). 243–285.
- Güldemann, Tom. 1999. Head-initial meets head-final: Nominal suffixes in eastern and southern Bantu from a historical perspective. *Studies in African Linguistics* 29. 49–91.
- Güldemann, Tom. 2004. Linear order as a basic morphosyntactic factor in Non-Khoe Khoisan. Paper presented at the the Syntax of the World's Languages conference in Leipzig, 8 August 2004.
- Güldemann, Tom. 2006. "Janus-headed" nominals: The morphosyntax of agreement in Taa. Paper presented at the International Conference "Rara & Rarisima – collecting and interpreting unusual characteristics of human languages", Leipzig, 29 March–1 April 2006.
- Güldemann, Tom & Ines Fiedler. 2019. Niger-Congo "noun classes" conflate gender with deriflection. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 95–145. Berlin: Language Science Press. DOI:10.5281/zenodo.3462762
- Haas, Mary R. 1940. Tunica. In *Handbook of American Indian Languages*, vol. 4, 1–143. New York: Augustin.
- Hammarström, Harald, Robert Forkel & Martin Haspelmath (eds.). 2018. *Glottolog 3.2*. Jena: Max Planck Institute for the Science of Human History. <http://glottolog.org/>.
- Handschuh, Corinna. 2018. Distinct marking of common and proper nouns in Oceanic (and beyond): Synchronic variation in form and function and historical implications. International Conference of Austronesian Linguistics, 17–20 July, 2018, Abstract book, 130–131. University of Antananarivo, Ambohitsaina.
- Harbour, Daniel. 2008. *Morphosemantic number: From Kiowa noun classes to UG number features* (Studies in Natural Language and Linguistic Theory 69). Dordrecht: Springer.
- Harrell, Frank E. 2001. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer.
- Haspelmath, Martin. 1999. Long distance agreement in Godoberi (Daghestanian) complement clauses. *Folia Linguistica* 33(2). 131–151.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.
- Haspelmath, Martin. 2013. Occurrence of nominal plurality. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*.

- Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://wals.info/chapter/34>.
- Haspelmath, Martin. 2018. Toward a new conceptual framework for comparing gender systems and some so-called classifier systems. Talk presented at Stockholm University, Department of Linguistics on April 13, 2018.
- Haugen, Einar. 1972. The ecology of language. In Anwar Dil (ed.), *The ecology of language: Essays by Einar Haugen*, 325–339. Stanford: Stanford University Press.
- Heath, Jeffrey. 1975. Some functional relationships in grammar. *Language* 51(1). 89–104.
- 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.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Heine, Bernd & Mechthild Reh. 1984. *Grammaticalization and reanalysis in African languages* (Kölner Beiträge zur Afrikanistik 1). Hamburg: Buske.
- Heine, Bernd & Rainer Vossen. 1983. On the origin of gender in Eastern Nilotic. In Rainer Vossen & Marianne Bechhaus-Gerst (eds.), *Nilotic studies: Proceedings of the international symposium on languages and history of the Nilotic peoples, Cologne, January 4-6, 1982*, 245–268. Cologne: Reimer.
- Hengeveld, Kees. 2012. Referential markers and agreement markers in Functional Discourse Grammar. *Language Sciences* 34(4). 468–479.
- Herrmann, Tania. 2005. Relative clauses in English dialects of the British Isles. In Bernd Kortmann, Tanja Herrmann, Lukas Pietsch & Susanne Wagner (eds.), *A comparative grammar of British English dialects: Agreement, gender, relative clauses*, 1–123. Berlin: Mouton de Gruyter.
- Himmelmann, Nikolaus. 2001. Articles. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, 831–841. Berlin: Mouton de Gruyter.
- Hockett, Charles F. 1958. *A course in modern linguistics*. New York: Macmillan.
- Hopper, Paul J. & Sandra A. Thompson. 1984. The discourse basis for lexical categories in universal grammar. *Language* 60(4). 703–752.
- Hudson, Joyce. 1985. *Grammatical and semantic aspects of Fitzroy Valley Kriol* (Work Papers of SIL-AAB, Series A, Vol. 8). Darwin: Summer Institute of Linguistics.

- Huntley, David. 1980. The evolution of genitive-accusative animate and personal nouns in Slavic dialects. In Jacek Fisiak (ed.), *Historical morphology* (Trends in Linguistics: Studies and Monographs 17), 189–212. The Hague: Mouton.
- Ishiyama, Osamu. 2008. *Diachronic perspectives on personal pronouns in Japanese*. State University of New York at Buffalo. (Doctoral dissertation).
- Jespersen, Otto. 1949. *A modern English grammar on historical principles, part II*. London: George Allen & Unwin Ltd.
- Johnston, Raymond Leslie. 1980. *Nakanai of New Britain. The grammar of an Oceanic language* (Pacific linguistics B70). Canberra: Australian National University.
- Jurafsky, Daniel. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72. 533–578.
- Karatsareas, Petros. 2009. The loss of grammatical gender in Cappadocian Greek. *Transactions of the Philological Society* 107. 196–230.
- Karatsareas, Petros. 2014. On the diachrony of gender in Asia Minor Greek: The development of semantic agreement in Pontic. *Language Sciences* 43. 77–101.
- Kayser, MSC., Alois. 1993. *Nauru grammar*. Edited by Karl H. Rensch. Yarralumla (Australia): Embassy of the Federal Republic of Germany.
- Kibrik, Andrej A. 2011. *Reference in discourse*. Oxford: Oxford University Press.
- Kilarski, Marcin. 2013. *Nominal classification: A history of its study from the classical period to the present*. Amsterdam: John Benjamins.
- Killian, Don. 2015. *Topics in Uduk phonology and morphosyntax*. University of Helsinki. (Doctoral 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:10.5281/zenodo.3462764
- Kirby, Simon, Hannah Cornish & Kenny Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31). 10681–10686. DOI:10.1073/pnas.0707835105
- Kohnen, Fr. Bernardo. 1933. *Shilluk grammar, with a little Shilluk-English dictionary*. Verona: Missioni Africane.
- Koptjevskaja-Tamm, Maria. 2005. Maria's ring of gold: Adnominal possession and non-anchoring relations in the European languages. In Ji-yung Kim, Yu Lander & Barbara H. Partee (eds.), *Possessives and beyond: Semantics and syntax*, 155–181. Amherst, MA: GLSA Publications.

- 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.
- Kuhn, Jonas & Louisa Sadler. 2007. Single conjunct agreement and the formal treatment of coordination in LFG. In Miriam Butt & Tracy Holloway King (eds.), *On-line Proceedings of the LFG2007 Conference*. <http://journals.linguisticsociety.org/proceedings/index.php/BLS/issue/view/120>.
- Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflections*. University of Leiden. (Doctoral dissertation). Utrecht: LOT.
- Kusters, Wouter. 2008. Complexity in linguistic theory, language learning and language change. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 3–22. Amsterdam: John Benjamins.
- Lass, Roger. 1990. How to do things with junk: Exaptation in language evolution. *Journal of Linguistics* 26. 79–102.
- Leeding, Velma J. 1989. *Anindilyakwa phonology and morphology*. University of Sydney. (Doctoral dissertation).
- Lehmann, Christian. 1982. Universal and typological aspects of agreement. In Hansjakob Seiler & Franz Josef Stachowiak (eds.), *Apprehension: Das sprachliche Erfassen von Gegenständen*. Vol. 2: *Universal and typological aspects of agreement: Die Techniken und ihre Zusammenhänge in Einzelsprachen*, 201–267. Tübingen: Narr.
- Lehmann, Christian. 1984. *Der Relativsatz. Typologie seiner Strukturen, Theorie seiner Funktionen, Kompendium seiner Grammatik* (Language Universals Series 3). Tübingen: Gunter Narr.
- 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:10.5281/zenodo.3462772
- Lindström, Eva. 1995. Animacy in interrogative pronouns. In Inger Moen, Hanne Gram Simonsen & Helge Lødrup (eds.), *Papers from the 16th Scandinavian Conference of Linguistics*, 307–315. Oslo: University of Oslo.
- Loughnane, Robyn. 2009. *A grammar of Oksapmin*. University of Melbourne. (Doctoral dissertation).
- Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLOS one* 5(1). e8559.

- Luraghi, Silvia. 2011. The origin of the Proto-Indo-European gender system: Typological considerations. *Folia Linguistica* 45(2). 435–464.
- Luraghi, Silvia. 2015. From non-canonical to canonical agreement. In Hans Amstutz, Andreas Dorn, Matthias Müller, Miriam Ronsdorf & Sami Uljas (eds.), *Fuzzy boundaries*, 71–88. Hamburg: Widmaier.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- Malkiel, Yakov. 1957–1958. Diachronic hypercharacterization in Romance. *Archivum Linguisticum* 9–10. 79–113, 1–36.
- Mathiot, Madeleine & Marjorie Roberts. 1978. Sex roles as revealed through referential gender in American English. In Madeleine Mathiot (ed.), *Boas, Sapir and Whorf revisited*, 1–47. The Hague: Mouton.
- McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2-3). 125–166.
- Meeuwis, Michael. 2013. Lingala. In Susanne Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The survey of pidgin and creole languages*. Vol. 3: *Lingala: Contact languages based on languages from Africa, Asia, Australia and the Americas*, 25–33. Oxford: Oxford University Press.
- Meillet, Antoine. 1897. *Recherches sur l'emploi du génitif-accusatif en vieux-slave* (Bulletin de l'Ecole des hautes études 115). Paris: Bouillon.
- Mellow, Greg. 2013. *A dictionary of Owa: A language of the Solomon Islands*. Berlin: De Gruyter Mouton.
- Merrifield, William R. 1959. Classification of Kiowa nouns. *International Journal of American Linguistics* 25(4). 269–271.
- Mettouchi, Amina. 2000. La “t” n'est-il qu'une marque de féminin en berbère (kabyle)? *Faits de Langue* 14. 217–225.
- Miestamo, Matti. 2008. Grammatical complexity in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (eds.), *Language complexity: Typology, contact, change*, 23–41. Amsterdam: John Benjamins.
- Miestamo, Matti, Kaius Sinnemäki & Fred Karlsson (eds.). 2008. *Language complexity: Typology, contact, change*. Amsterdam: John Benjamins.
- Mithun, Marianne. 1986. The convergence of noun classification systems. In Collette Craig (ed.), *Noun classes and categorization. Proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983*, 379–397. Amsterdam: John Benjamins.
- Mithun, Marianne. 1999. *The languages of native North America*. Cambridge: Cambridge University Press.
- Mithun, Marianne. 2014. Gender and culture. In Greville Corbett (ed.), *The expression of gender*, 132–160. Berlin: Mouton de Gruyter.

- Moravcsik, Edith. 1978. Agreement. In Joseph Greenberg (ed.), *Universals of human language*. Vol. 4: *Agreement: Syntax*, 331–374. Standord: Stanford University Press.
- Moravcsik, Edith. 1994. Parts and wholes in definite article constructions. Paper presented at the Pre-inaugural meeting of ALT, Konstanz, November 30.
- Moscatti, Sabatino. 1964. *An introduction to the comparative grammar of the Semitic languages*. Wiesbaden: Otto Harrasowitz.
- Mous, Maarten. 2008. Number as an exponent of gender in Cushitic. In Zygmunt Frajzyngier & Erin Shay (eds.), *Interaction of morphology and syntax: Case studies in Afroasiatic*, 137–160. Amsterdam: John Benjamins.
- Nagaraja, Keralapura Shreenivasaiah. 1996. The status of Lyngngam. *Mon-Khmer Studies: A Journal of Southeast Asian Linguistics and Languages* 26. 37–50.
- Nagaraja, Keralapura Shreenivasaiah, Paul Sidwell & Simon J. Greenhill. 2013. A lexicostatistical study of the Khasian languages: Khasi, Pnar, Lyngngam, and War. *Mon-Khmer Studies* 42. 1–11.
- Nau, Nicole. 1999. Was schlägt der Kasus? Zu Paradigmen und Formengebrauch von Interrogativpronomina. *Sprachtypologie und Universalienforschung* 52. 130–150.
- Nau, Nicole. 2011. Declension classes in Latvian and Latgalian: Morphemics vs. morphophonology. *Baltic Linguistics* 2. 141–177.
- Newman, Paul. 1979. Explaining Hausa feminines. *Studies in African Linguistics* 10(2). 197–226.
- 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.
- Nichols, Johanna. 2009. Linguistic complexity: A comprehensive definition and survey. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable*, 110–125. Oxford: Oxford University Press.
- 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:10.5281/zenodo.3462760
- Olsson, Bruno. 2017. *The Coastal Marind language*. Singapore: Nanyang Technological University. (Doctoral dissertation).
- 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:10.5281/zenodo.3462768
- Oomen, Antoinette. 1981. Gender and plurality in Rendille. *Afroasiatic linguistics* 8(1). 35–75.
- Orkaydo, Ongaye Oda. 2013. *A grammar of Konso*. Utrecht: LOT: Netherlands Graduate School of Linguistics. (Doctoral dissertation).
- Parker, Enid M. & Richard J. Hayward. 1985. *An Afar-English-French dictionary (with grammatical notes in English)*. London: School of Oriental & African Studies (SOAS). 308.
- Passer, Matthias Benjamin. 2016a. (What) do verb classifiers classify? *Lingua* 174(1). 16–44.
- Passer, Matthias Benjamin. 2016b. *The typology and diachrony of nominal classification*. Utrecht: University of Amsterdam. (Doctoral dissertation).
- Pawley, Andrew. 2004. Using *he* and *she* for inanimate referents in English: Questions of grammar and world view. In Nick J. Enfield (ed.), *Ethnosyntax, explorations in grammar and culture*, 110–137. Oxford: Oxford University Press.
- Payne, Doris L. 1985. *Aspects of the grammar of Yagua: A typological perspective*. University of California, Los Angeles. (Doctoral dissertation).
- Payne, Doris L. 1986. Noun classification in Yagua. In Colette G. Craig (ed.), *Noun classes and categorization. Proceedings of a symposium on categorization and noun classification, Eugene, Oregon, October 1983*, 113–131. Amsterdam: John Benjamins.
- Payne, Doris L. 1998. Maasai gender in typological perspective. *Studies in African Linguistics* 27(2). 159–175.
- Penchoen, Thomas. 1973. *Etude syntaxique d'un parler berbère (Ait-Frah de l'Aurès)*. Napoli: Centro di studi magrebini.
- Petrollino, Sara. 2016. *A grammar of Hamar: A South Omotic language of Ethiopia*. Cologne: Köppe.
- Plank, Frans. 2015. Genus im Gastgewerbe. Mit einem Nachtrag: Der/das Zwiebfisch und die Tiefe der Jahre. Manuscript.
- Plaster, Keith & Maria Polinsky. 2007. Women are not dangerous things: Gender and categorization. *Harvard Working Papers in Linguistics* 12. 1–44.
- Polinsky, Maria & Bernard Comrie. 1999. Agreement in Tsez. *Folia Linguistica* 33(2). 109–130.
- Polinsky, Maria & Ezra Van Everbroeck. 2003. Development of gender classifications: Modeling the historical change from Latin to French. *Language* 79. 356–390.

- Rabel-Heymann, Lili. 1977. Gender in Khasi nouns. *Mon-Khmer Studies* 6. 247–272.
- Reid, Nicholas. 1990. *Ngan'gityemerri: A language of the Daly River region, Northern Territory of Australia*. Canberra: Australian National University. (Doctoral dissertation).
- Reid, Nicholas. 1997. Class and classifier in Ngan'gityemerri. In Mark Harvey & Nicholas Reid (eds.), *Nominal classification in Aboriginal Australia*, 165–225. Amsterdam: John Benjamins.
- Ring, Hiram. 2015. *A grammar of Pnar*. Singapore: Nanyang Technological University. (Doctoral dissertation).
- Rissanen, Matti. 1997. The pronominalization of *one*. In Matti Rissanen, Merja Kytö & Kirsi Heikkonen (eds.), *Grammaticalization at work. Studies of long-term developments in English*, 87–143. Berlin: Mouton de Gruyter.
- Rizzi, Luigi. 1982. A restructuring rule. In Luigi Rizzi (ed.), *Issues on Italian syntax*, 1–48. Dordrecht: Foris.
- Rudzīte, Marta. 1964. *Latviešu izlokšņu teksti [Texts in Latvian dialects]*. Rīgā: Latvijas Valsts universitāte. Reprinted in Rudzīte, Marta 2005. *Darbi latviešu dialektoloģijā*, 173–251. Rīga: LU Akadēmiskais apgāds.
- Russell, Bertrand. 1905. On denoting. *Mind, New Series* 14. 479–493.
- Sagna, Serge. 2011. Semantic categorizations in the Gújjolaay Eegimaa collectives and distributives. In Oliver Bond, Peter K. Austin, David Nathan & Marten Lutz (eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory 3*, 237–246. London: Department of Linguistics, SOAS.
- Sampson, Geoffrey, David Gil & Peter Trudgill (eds.). 2009. *Language complexity as an evolving variable*. Oxford: Oxford University Press.
- Santos Sachot, Rosine. 1996. *Le mey*. Paris: Université de la Sorbonne Nouvelle. (Doctoral dissertation).
- Sasse, Hans-Jürgen. 1993. Syntactic categories and subcategories. In Joachim Jakobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: Ein internationales Handbuch zeitgenössischer Forschung / An international handbook of contemporary research*, 646–686. Berlin: Walter de Gruyter.
- Schachter, Paul & Fe T. Otanes. 1972. *Tagalog reference grammar*. Berkeley: University of California Press. Reprinted in 1983.
- Schuh, Russell G. 1998. *A grammar of Miya* (University of California Publications in Linguistics 130). Berkeley & Los Angeles: University of California Press. 414.
- Schuster-Šewc, Heinz. 1976. *Gramatika hornjoserbskeje řeče [Grammar of Upper Sorbian]*. Vol. 2: Syntaksa. Bautzen: Domowina.

- Seifart, Frank. 2005. *The structure and use of shape-based noun classes in Miraña*. Nijmegen: Radboud Universiteit Nijmegen. (Doctoral dissertation).
- Seifart, Frank. 2007. The prehistory of nominal classification in Witotoan languages. *International Journal of American Linguistics* 73(4). 411–445.
- Seifart, Frank. 2010. Nominal classification. *Language and Linguistics Compass* 4(8). 719–736.
- Seifart, Frank. 2018. The semantic reduction of the noun universe and the diachrony of nominal classification. In William B. McGregor & Søren Wichmann (eds.), *The diachrony of classification systems*, 9–32. Amsterdam: John Benjamins.
- Senft, Gunter. 1985. Klassifikationspartikel im Kilivila: Glossen zu ihrer morphologischen Rolle, ihrem Inventar und ihrer Funktion in Satz und Diskurs. *Linguistische Berichte* 99. 373–393.
- Senft, Gunter. 1986. *Kilivila: The language of the Trobriand islanders*. Berlin: Mouton de Gruyter.
- Senft, Gunter. 1993. A grammaticalization hypothesis on the origin of Kilivila classificatory particles. *Sprachtypologie und Universalieforschung* 46(2). 100–112.
- Seuren, Pieter A. M. 2009. Donkey sentences. In Keith Allan (ed.), *Concise encyclopedia of semantics*, 269–271. Oxford: Elsevier.
- Siemund, Peter. 2008. *Pronominal gender in English: A study of English varieties from a cross-linguistic perspective*. New York: Routledge.
- Siemund, Peter & Florian Dolberg. 2011. From lexical to referential gender: An analysis of gender change in medieval English based on two historical documents. *Folia Linguistica* 45(2). 489–534. DOI 10.1515/flin.2011.018.
- Silverstein, Michael. 1976. Hierarchy of features and ergativity. In Robert M. W. Dixon (ed.), *Grammatical categories in Australian languages*, 112–190. New Jersey: Humanities Press.
- Singer, Ruth. 2011. Typologising idiomaticity: Noun-verb idioms and their relations. *Linguistic Typology* 15. 625–569.
- Singer, Ruth. 2012. Do nominal classifiers mediate selectional restrictions? An investigation of the function of semantically-based nominal classifiers in Mawng (Iwaidjan, Australian). *Linguistics* 50. 995–990.
- Singer, Ruth. 2018. Beyond the classifier/gender dichotomy: The role of flexibility in a more integrated typology of nominal classification. In Sebastian Fedden, Jenny Audring & Greville G. Corbett (eds.), *Non-canonical gender systems*, 100–128. Oxford: Oxford University Press.

- Sinnemäki, Kaius. 2014a. Cognitive processing, language typology, and variation. *WIREs Cognitive Science* 5(4). 477–487.
- Sinnemäki, Kaius. 2014b. 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. 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:10.5281/zenodo.3462782
- Sinnemäki, Kaius & Francesca Di Garbo. 2018. Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology* 9. DOI:10.3389/fpsyg.2018.01141
- Small, Priscilla C. 1990. A syntactic sketch of Coatzospan Mixtec. In C. Henry Bradley & Barbara E. Hollenbach (eds.), *Studies in the syntax of Mixtecan languages 2* (Summer Institute of Linguistics and the University of Texas at Arlington Publications in Linguistics 90), 261–479. Dallas: Summer Institute of Linguistics & the University of Texas at Arlington.
- 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.
- Smith, Kenny & Elizabeth Wonnacott. 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116. 444–449.
- Spagnolo, Lorenzo M. 1933. *Bari grammar*. Verona: Missioni Africane.
- Stolz, Thomas. 2012. Survival in a niche. On gender-copy in Chamorro (and sundry languages). In Martine Vanhove, Thomas Stolz, Aina Urdze & Hitomi Otsuka (eds.), *Morphologies in contact*, 93–140. Munich: Akademie-Verlag.
- Straus, Anne Terry & Robert Brightman. 1982. The implacable raspberry. *Papers in Linguistics* 15. 97–137. DOI:10.1080/08351818209370564
- Sutton, Logan. 2014. *Kiowa-Tanoan: A synchronic and diachronic study*. Albuquerque: University of New Mexico. (Doctoral dissertation).
- 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:10.5281/zenodo.3462770

- Swanton, John R. 1921. The Tunica language. *International Journal of American Linguistics* 2(1/2). 1–39.
- Tasmowski-De Ryck, Liliane & Paul Verluyten. 1982. Linguistic control of pronouns. *Journal of Semantics* 1. 323–346.
- Teleman, Ulf, Staffan Hellberg & Erik Andersson. 1999. *Svenska Akademiens grammatik [Grammar of the Swedish Academy]*. Vol. 2: Ord. Stockholm: Nordstedts.
- Thiesen, Wesley & David Weber. 2012. *A grammar of Bora with special attention to tone* (SIL International Publications in Linguistics 148). Dallas: SIL International.
- Thomason, Lucy Grey. 2003. *The proximate and obviative contrast in Meskwaki*. University of Texas at Austin. (Doctoral dissertation).
- Traill, Anthony. 1994. *A !Xóõ dictionary*. Cologne: Rüdiger Köppe Verlag.
- Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. New York: Oxford University Press.
- Tsegaye, Mulugeta. 2017. *Plural gender: Behavior evidence for plural as a value of Cushitic gender with reference to Konso*. Leiden: Leiden University. (Doctoral dissertation).
- Tucker, A. N. & J. Tompo Ole Mpaayei. 1955. *A Maasai grammar with vocabulary* (Publications of the African Institute, Leyden 2). London: Longmans, Green & Co.
- Universals Archive. n.d. *Konstanz universals archive*. <https://typo.uni-konstanz.de/archive/>. Compiled by Frans Plank and collaborators.
- Usher, Timothy & Edgar Suter. 2015. The Anim languages of Southern New Guinea. *Oceanic Linguistics* 54(1). 110–142.
- Vajda, Edward J. 2003. Ket verb structure in typological perspective. *STUF - Language Typology and Universals* 56(1-2). 55–92. DOI:10.1524/stuf.2003.56.12.55
- van den Heuvel, Wilco. 2006. *Biak: Description of an Austronesian language of Papua*. Vrije Universiteit Amsterdam. (Doctoral dissertation).
- van der Meer, Suzanne. 2015. *The distribution of P-nouns in Cushitic*. Leiden University. (MA thesis).
- Van Valin, Robert D. 1977. *Aspects of Lakhota syntax: A study of Lakhota (Teton Dakota) syntax and its implications for universal grammar*. University of California, Berkeley. (Doctoral dissertation).
- van Driem, Georg. 1987. *A grammar of Limbu*. Berlin: Mouton de Gruyter.
- van Rijn, Marlou. 2016. The grammaticalization of possessive person marking: A typological approach. *Transaction of the Philological Society* 114(2). 233–276.

- Verhaar, John W. M. 1995. *Toward a reference grammar of Tok Pisin: An experiment in corpus linguistics* (Oceanic Linguistics Special Publication 26). Honolulu: University of Hawai'i Press.
- Verkerk, Annemarie. 2014. Change in Bantu noun class systems. Poster presented at the Nijmegen Lectures 2014 (Nijmegen, 27–29 January 2014).
- Wälchli, Bernhard. 2017. The incomplete story of feminine gender loss in Northwestern Latvian dialects. *Baltic Linguistics* 8. 143–214.
- 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.
- 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:10.5281/zenodo.3462780
- Watkins, Laurel J. & Parker McKenzie. 1984. *A grammar of Kiowa*. Lincoln: University of Nebraska Press.
- Weidert, Alfons. 1975. *I Tkong Amwi: Deskriptive Analyse eines Wardialekts des Khasi* (Neuindidische Studien). Wiesbaden: Otto Harrassowitz.
- Widmark, Gun. 1966. Den inkongruenta neutrala predikatsfyllnaden och dess plats i svenskans genussystem [The non-agreeing neuter complement and its place in the current Swedish gender system]. *Nysvenska studier* 46. 91–135.
- Wolfart, H. Christoph. 1973. *Plains Cree: A grammatical study* (Transactions of the American Philosophical Society, New Series 63.5). Philadelphia: American Philosophical Society.
- Wonderly, William L., Paul L. Kirk & Lorna F. Gibson. 1954. Number in Kiowa: Nouns, demonstratives, and adjectives. *International Journal of American Linguistics* 20. 1–7.
- Wunderlich, Dieter. 1994. Towards a lexicon-based theory of agreement. *Theoretical Linguistics* 20. 1–35.
- Zaliznjak, Andrey A. 1977. *Grammatičeskij slovar' russkogo jazyka: Slovoizmenenie* [Grammatical dictionary of Russian: Inflection]. Moskva: Russkij Jazyk.
- Zipf, George Kingsley. 1935. *The psycho-biology of language. An introduction to dynamic philology*. Boston: Houghton Mifflin.

Appendix: List of topics with short definitions and where these are treated in the chapter

Absolute complexity: complexity as an objective property of grammatical domains (§2.1).

Absolute condition: also obligatory condition: condition that always determines a certain choice of agreement value (§7.5).

Adjacency: controller and target or target and controller follow each other immediately. A possible specific relationship in agreement (§7.2).

Adnominal use (of nominal morphology): marker on an adnominal modifier or dependent in an NP with a noun head or NP marker in a NP with a noun head (§4.1).

Adnominal modifier: modifier in an NP with a head noun, such as attributive adjective, relative clause with a head noun, attributive demonstrative and attributive numeral (§4.1).

Agentivity: semantic connection between animacy and inanimate objects, responsible for the fact that agentive nouns are more likely to take an animate gender (§3.3 (i)).

Agreement: an asymmetric specific relation between a controller and a target involving displaced information. Agreement is syntactic to the extent that it involves words or groups of words as targets and controllers, but the relation between controller and target can be semantic (as in inter-sentential agreement). Controllers can be latent (contextual, semantic) (§7.1).

Agreement Hierarchy: more distal controllers are more likely to trigger semantic agreement along a hierarchy attributive < predicate < relative pronoun < personal pronoun (Corbett 1991: 226) (§3.6).

Anaphor, pl. anaphora: linguistic element that is lacking clear independent reference and picks up reference through connection with another element.

Animacy distinctions: the linguistic encoding of the ontological difference between living and non-living beings.

Animacy hierarchy: certain patterns of language structure (e.g., plural marking, differential object marking) are more likely to emerge/be synchronically restricted to humans and or highly animate entities only. Based on these

effects, types of entities can be arranged on a hierarchy of degree of animacy: speaker > addressee > 3rd person > kinship terms > other humans > “higher” animals > “lower” animals > discrete inanimates > nondiscrete inanimates (Smith-Stark 1974; Corbett 2000; Haspelmath 2013).

Apposition: two nominal constituents in the same case role and not in a predicative relationship and not in a relationship of subordination (none of the two is the head of the other one) (§4.3).

Areal pattern: distribution of linguistic properties across languages in a geographical area that cannot be explained by obvious genealogical relation of languages (§11.1).

Articulatory plurality: lexical plurality expressed with double-hand signs in sign languages (§9.1).

Associated gender: noun receiving its gender through a link with another noun (§3.4).

Augmentative: grammatical construction that, in its basic meaning, expresses that a given entity is bigger than its standard size (§5.2).

Canonical Approach: theoretical and methodological approach to the typological study of morphosyntactic features developed by Greville Corbett. The approach is based on the idea that, for every morphosyntactic phenomenon, there exists a space of crosslinguistic variation and that attested language-specific systems are situated in this space in ways that more or less correspond to a certain identified base of comparison (Audring & Fedden 2018: 2) (§1).

Case: marker of grammatical relation or oblique semantic role, often cumulating with gender (§8.2).

Classifiers: cover term for numeral classifiers, noun classifiers, and possessive classifiers, and some further minor types of classifiers (§1).

Co-conceptuality: a specific relationship in agreement where controller and target express identity of concept (but not identity of reference) (§7.2).

Co-evolution: a set of more than one diachronic change that are at least partly dependent on each other (§10.2).

- Complex controller:** the agreement controller consists of several words (§7.3).
- Complex:** (i) non-trivial in structure, so that an exhaustive description cannot be short. But also (ii) consisting of several elements and (iii) heterogeneous, consisting of various, but related phenomena (§1).
- Complex target:** the agreement target consists of several words. These constitute a formal group (§7.4).
- Compound gender:** the gender of a compound is different from the gender of its head (§6.2).
- Concurrent gender systems:** two or more than two gender systems that are largely independent of each other within the same language (§1).
- Condition:** factor provoking the choice of an agreement value, can be absolute or relative (§7.5).
- Contrastive focus:** emphasis of a choice of argument as opposed to another or other possible choices, induces transparency in gender-marked anaphoric pronouns by activating the descriptive content of gender (§3.6).
- Controller:** formal or contextual element triggering the choice of a marker of a grammatical category (such as gender or number) (§7.1).
- Coreferentiality:** a specific relationship in agreement where controller and target have identity of reference (§7.2).
- Covert marking of gender:** extent to which nouns lack formal gender assignment (§6.4).
- Cumulation:** expression of two or more grammatical categories in the same morpheme (§8).
- Decategorialization of nouns:** nouns losing some of their prototypical properties, notably when used in non-referential contexts (e.g., predicatively) (§7.6).
- Declension class:** morphological paradigm (according to number, case, and/or any other nominal grammatical category) characterizing a subset of nouns (§1, §6.4).
- Default:** rest category for gender assignment, usually thought of as last resort (§6.5).

- Derivational gender:** gender in nominal targets expressed on nouns by derivational morphology (§3.4, §7.6).
- Description length:** from an information theory perspective, one of the ways of measuring system complexity. The longer its description, the more complex the system (§2.1).
- Descriptive complexity:** (or Kolmogorov complexity), the information required to describe a system (the longer the description the more complex the system) (§2.1).
- Differential case marking:** a grammatical relation is indicated by different case forms or appositions, often depending on animacy and/or definiteness (§8.2).
- Differential object marking:** the grammatical relation object is indicated by different case forms or appositions, often depending on animacy and/or definiteness (§3.5).
- Diminutive:** grammatical construction that, in its basic meaning, expresses that a given entity is smaller than its standard size. Additional meanings associated with diminutive constructions are: affection, partitive, female (see Jurafsky 1996 for a full list) (§5.2).
- Displacement (of information):** the word or context triggering the choice of a grammatical value of a marker does not originate in the word on which the category is marked (§6.4).
- Dynamic approach:** viewing a set of related phenomena as something that can emerge, evolve and disappear in accordance with certain diachronic pathways of development and assuming that these developments are crucial for the understanding of the phenomena (§1).
- Ecology of languages:** the interaction between any given language and its natural and/or social environment (Haugen 1972) (§11.2).
- Feature:** a fully paradigmaticized grammatical category type expressed by systematic morphological marking. Typical examples of features are: gender, number, case, person, and tense (Corbett 2012) (§7.5).
- Formal assignment:** morphological and/or phonological gender assignment and opposed to semantic gender assignment (§5.1, §6.4).

Formal group: several words together constituting a syntactic unit (can but need not be a constituent; Croft 2001: 190) (§7.1).

Gender: gender is a grammatical category type with a semantic core of animacy and/or sex reflecting classes of referents, which have a propensity to turn into classes of noun lexemes. It is overtly marked on noun-associated forms. It typically exhibits cumulative expression with number, case, and/or person. Gender is organized in the form of systems (§1).

Gender assignment: rationale determining the gender of a noun (can be semantic or formal) (§5).

Gender recategorization: the phenomenon whereby gender assignment is not fixed but subject to variation based on reference construal. Synonymous with: *flexible/manipulable gender assignment*. But also used for reconceptualization of same referent in discourse (§3.7).

Gender resolution: the gender of a complex controller is determined by means of interaction between the genders of at least two of its parts (§7.3).

Gendered clause: subordinate clause (often an independent relative clause) bearing a gender marker (§4.3).

Gender value: one gender from the set of genders in a gender system (§1).

Grammatical anaphor: anaphor intermediate between pronoun (third person pronoun) and noun (noun in anaphoric function like *that man*) (§4.3).

Headedness reversal: a semantic modifier or dependent of a phrase is its formal head (§4.3).

Hierarchical patterning: organization of the structure of a grammatical category according to a hierarchy (§3).

Hybrid noun: noun that can trigger two or more different gender values (but often only one of them is lexical gender) (§3.6).

Hypercharacterization: diachronic process whereby a marker is added that overtly indicates a category that the element already had before (§6.4).

Idiomatization of gender: a particular use of gender is restricted to idioms or an idiom (§7.7).

Independent modifier: modifier in an NP without a head noun, such as free relative clause, pronominal demonstratives and pronominal numerals (§4.1).

Independent nominal morphology: grammatical marking in an NP without a head noun (§4.1).

Indexation: an index is a bound or free grammatical marker – prototypically a marker of person – that denotes the argument itself. One argument can be marked several times by different indexes, which are then in a relationship of coreference (§7.1).

Individuation hierarchy: version of the animacy hierarchy subdividing inanimates into tangible objects, abstracts and mass nouns (Sasse 1993) (§3.2).

Information transfer chain: displacement of information in agreement in several steps, e.g., from gender assignment to noun lexeme to word-form to complex controller to complex target to word within target to gender marker realized on that word (§7.1).

Inherited gender: the possessor determines the gender of a noun or NP (§3.4).

Inter-sentential agreement: controller and target in agreement are or can be in different sentences (§7.1).

Intra-sentential agreement: controller and target in agreement occur within the same sentence (§7.1).

Inventory complexity: the number of distinctions in a grammatical system (§2.1).

Latent controller: a contextual controller that is not realized in syntax (§7.1).

Lexical gender: classes of noun lexemes distinguished on noun-associated forms (§3.1).

Local agreement: agreement within the noun phrase (§4.1).

Mature phenomenon: a phenomenon with a non-trivial prehistory (Dahl 2004: 2) (§1).

Many-to-one gender assignment: the same gender value is the outcome of several assignment rules (§6.2).

- Morphological gender assignment:** the gender value of a controller is determined by some of its inherent morphological properties (e.g., its declension class) (§5.1, §6.4).
- Neutral gender:** agreement form used for agreement with non-noun controllers, such as infinitive phrases, clauses, interjections and quoted phrases (§6.5).
- Nomifier:** cover term for gender and classifiers (§1).
- Nominal gender targets:** nouns or noun phrases that are gender targets (typically decategorialized nouns) (§7.6).
- Nominal morphology:** cover term for non-lexical markers within the noun phrase (§4.1).
- Non-noun controllers:** a controller in agreement that is not a noun, see neutral gender (§6.5).
- Noun-associated form:** an adnominal modifier (article, demonstrative, adjective, or numeral), or a verbal argument index, or an anaphoric pronoun (§1).
- Noun class:** same as gender, but emphasizing classes of noun lexemes (§1).
- Noun incorporation:** compound of a noun (usually in object function) and its verbal head. Has classifying potential to the extent the incorporated nouns are hyperonymic (§7.6).
- Nominal target:** agreement is imposed on a noun or noun phrase (§7.6).
- Number:** grammatical category marking number of referents (singular, plural, dual, non-singular etc.), frequently cumulating with gender (§8.1).
- One-to-one gender assignment:** every gender assignment rule applies to another gender value (§6.2).
- Opaque gender assignment:** non-formal gender assignment that is not general but characterized by numerous exceptions (§5.1, §6.3).
- Overt marking of gender:** extent to which nouns exhibit formal gender assignment (§5.1, §6.4).
- Person:** grammatical category indicating whether or not a referent is a speech act participant and which one (speaker or addressee), marked in free or bound personal pronouns (also called indexes). It may be sensitive to honorific distinctions. Person frequently cumulates with gender (§8.2).

Person name marker: marker indicating that an element is a name of a person, can be a general person name marker or distinguish male and female names; also called proprial article, but not all person name markers are articles and “proprial” and “proper name” does not specify that the markers tend to be dedicated to person names rather than place names (§3.5).

Phonological gender assignment: the gender value of a controller is determined by some of its inherent phonological properties (§5.1, §6.4).

Plurale tantum, pl. Pluralia tantum: literally, noun that only exists in the plural, but more broadly noun exhibiting lexical plurality (plural is a lexicalized property of a noun) (§9).

Principle of Contrast: captures the observation that in systems with opaque gender assignment nouns in a semantic field preferably have a dominant gender, but some salient nouns in the field tend to contrast with them and take an opposite gender (§6.3).

Principle of Fewer Distinctions (also Principle of Economy): measure of inventory complexity stating that the fewer distinctions made the less complex the domain (§2.1).

Principle of One-Meaning–One-Form (also Principle of Transparency): measure of transparency whereby the less complex grammatical phenomenon is one where there is a one-to-one correspondence between meaning and form (§2.1).

Principle of Independence: measure of complexity whereby the less complex grammatical domain/pattern of encoding is the one that is NOT dependent on another grammatical domain/pattern of encoding (§2.1).

Power: semantic connection between animacy and inanimate objects concerning objects endowed with some inherent potential of agency (§3.3 (iv)).

Pronominal articles: use of personal pronouns (third person and occasionally others) with noun phrases often with some restriction to referential, specific or animate (§3.5).

Pronominal gender systems: gender systems with pronouns as the only agreement target (§3.2).

- Purview:** semantic connection between animacy and inanimate objects concerning objects that belong to or relate to a human or animate referent, are perceived as human or animate in size, shape, or function, or are spoken about by humans (Gerdts 2013) (§3.3 (iii)).
- Reconceptualization of referents (also called recategorization):** switch of gender in a sequence of coreferential expressions due to association with different gender controllers with different gender (§3.7).
- Referent-based gender:** Dahl's (2000a) "referential gender". Classes of referents distinguished on noun-associated forms (§3.1).
- Relative condition, also optional condition:** factor that favors a certain optional choice of agreement value (§7.5).
- Relevance:** the meaning of an element is relevant to another to the extent its semantic content interferes with the meaning of the other element (Bybee 1985) (§6.4).
- Repeater:** classifier with the same form as the noun classified (§4.3).
- Sandhi:** phonological processes across word boundaries (§6.4).
- Saliency:** semantic connection between animacy and inanimate objects based on the fact that discourse prominence of referents can be viewed as an aspect of animacy (§3.3 (ii)).
- Semantic agreement:** agreement with a referent-based controller (i.e., referent-based gender), often not local, can be inter-sentential. Follows the Agreement Hierarchy (§3.6).
- Semantic gender assignment:** the gender value of a controller is determined by some of its semantic properties, can be lexical gender or referent-based gender (§5.1).
- Specific relationship in agreement:** property rendering the relation between controller and target unequivocal, e.g. coreference, co-conceptualization, certain kind of syntactic dependency, adjacency (§7.2).
- System:** minimally an opposition of at least two markers; however, often much more organized. If the notion is restricted to the more complex cases: highly organized language-specific complexes with both paradigmatic and

syntagmatic components that play an important role in the architecture of grammar (however that architecture is modeled). Systems must constantly and actively be dealt with in production and comprehension, which presupposes a high degree of adaptability to previously non-encountered discourse contexts (§1).

System evolution: cover term for all kinds of changes in the structure of systems and in particular including changes that entail the emergence of a system (§10).

Target: element or group of elements on which an agreement marker is realized (§7.1).

Target-controlled gender: the target restricts the choice of the gender value or contributes to the choice of gender value in another way (at the same time as there is displacement) (§7.7).

Unification: assumption that features of elements in an agreement relation are combined, which results in a symmetric interpretation of agreement. With unification, certain kinds of feature mismatches are tolerated and can be accounted for. Our approach does not provide feature mismatches, but accounts for the relevant phenomena with information transfer chains with several chain links (§7.1).

Uniqueness: a set of referents with a single member. Since animates are often unique (especially when referred to by person names), uniqueness is a semantic connection between animacy and inanimate objects (§3.3 (iv)).