Gender marking in the first-person singular: A case of paradigm (in)consistency*

THOMAS BERG

University of Hamburg

Since the sex of the speaker is normally as obvious as can be, there is no point in coding first-person singular gender – or so it may seem. This typological study examines the extent of sex-based gender marking in personal pronouns, possessive determiners, predicative adjectives, and verbs across first-, second-, and third-person singular. A worldwide perusal of grammars in addition to data elicitation yields a total of 115 languages with first-person gender. The paradigms of pronouns and possessives are found to be highly inconsistent, whereas those of verbs show a tendency towards consistency. Gender marking on adjectives is fully consistent. The likelihood of first-person gender is increased by a general sensitivity to gender and a dedicated gender morpheme. A distinction is made between pronouns and possessives as referential units and gender markers on verbs and adjectives as grammatical units. By their very nature, referential markers are sensitive to the contingencies of the extralinguistic world and subject to communicative constraints such as redundancy and economy. They therefore end up being organized in inconsistent paradigms. By

---

527

contrast, grammatical units are largely untouched by these extraneous influences and may therefore develop consistent paradigms.

KEYWORDS: gender, inflection, paradigm, person, pronoun

1. INTRODUCTION

For reasons of efficiency and economy, speakers must decide which information to put into words and which information to leave unexpressed. This decision harks back to Grice's (1975) Maxim of Quantity which reads: 'Do not make your contribution more informative than is required'. What criteria assist speakers in 'separating the wheat from the chaff'? A fairly large number of contrasts come to mind, including newsworthy vs. unnewsworthy, default vs. non-default, known vs. unknown to the listener, self-evident vs. non-self-evident, inferable vs. non-inferable, etc. These contrasts do not only play a role at the pragmatic level envisaged by Grice, but they are also reflected in language structure in terms of what is grammatically coded and what is not. As is well-known, languages are typically asymmetrically structured such that they mark non-default cases but zero-mark default cases (e.g. Haiman 1983; Mayerthaler 1988; Croft 2003). Pertinent examples abound. Singulars, affirmatives, and present tenses are usually unmarked, whereas plurals, negatives, and past tenses are usually marked. In all these categories, the unmarked cases are more predictable or more basic than the marked cases. It is worth noting that asymmetrical coding is not a logical necessity. Languages could certainly mark singulars and present tenses, as indeed some of them do. Rather, the general design principle appears to be to avoid overt marking when the listener can be reasonably expected to be able to reconstruct and fill in the relevant details.

Information may be inferable for a variety of linguistic and extralinguistic reasons. A particularly obvious source is the communicative context in which a speech activity is embedded. This encompasses speaker and listener attributes, the relationship between the interlocutors, the purpose of the encounter, and sundry other variables. One such factor, which is as obvious as any one factor can conceivably be, is the focus of the present study, viz the sex of the speaker. In face-to-face interaction, it is almost always easy to identify the sex of the interactants. No matter how the sex roles are distributed, listeners may derive this information from the physique, the voice, and many culture-specific variables such as clothes and gestures. Even in the absence of visual contact, interlocutors are not usually in doubt about the sex of the other. In a nutshell, the sex of the speaker is self-evident or, as Aikhenvald & Dixon (1998: 66) put it, 'communicatively redundant'.[1]

When we say that the sex of the speaker is self-evident, we conceive of sex in biological (and social) terms. In this sense, the sex of the speaker is certainly

---

[1] The notion of communicative redundancy does not imply that first-person gender has no functional motivation whatsoever. Redundancy has repeatedly been argued to serve as an aid in language learning and processing (e.g. Audring 2014 and Tal & Arnon 2022). On the function of (nominal) gender, see e.g. Dye, Milin, Futrell & Ramscar (2017) and Rogers & Gries (2022).

communicatively redundant. However, this view of language is a little too narrow. One of the many functions of language is to communicate one's identity. Since biological sex and our social construction thereof are an integral part of our identity, we may wish to make a point of it on certain occasions. We may approach topics from a particular perspective such as that of a woman or a man. As these two perspectives need not be identical, speakers may feel the need to resort to sex-specific marking. From this pragmatic angle, the sex of the speaker is NOT 'communicatively redundant'.

Let us label these opposing views the Redundancy Hypothesis and the Identity Hypothesis. A notable difference between them is that the Redundancy Hypothesis is listener-oriented, whereas the Identity Hypothesis is speaker-oriented. The sex of the speaker is obvious to the listener (as it is to the speaker him- or herself) while the notion of identity refers to the speaker's self-image. These different sources might explain why two hypotheses have to be entertained.

The two hypotheses make conflicting predictions about gender marking. The Redundancy Hypothesis predicts that the languages of the world leave the sex of the speaker unexpressed. In contrast, the Identity Hypothesis predicts the occurrence of gender marking in the first person. It is difficult to make this prediction more precise. In light of the Redundancy Hypothesis, it would be folly to expect all languages to code the sex of the speaker. The Identity Hypothesis gains credence if a certain number of languages with first-person gender marking are found. Such an empirical situation would not necessarily falsify the Redundancy Hypothesis but show instead that it is insufficient and that it has to be supplemented by the Identity Hypothesis.

According to Siewierska (2004: 105), the Redundancy Hypothesis makes different predictions for first and second vs. third person. Since she regards both first- and second-person gender as redundant, she does not predict a difference between them. However, the sex of the speaker and that of the listener are not equally obvious. While speakers are not normally in doubt about the sex of their interlocutors, they can identify the sex of others less reliably than their own. However minor this difference in reliability may be, language structure might be sensitive to it by marking second-person gender more often than first-person gender. In view of the fact that third-person referents are defined by their absence in a given communicative act, it is only to be expected that third-person marking occurs more frequently than second- (and first-) person marking. Hence, first-person gender marking is predicted to be less likely than second-person marking, which in turn is predicted to be less likely than third-person marking. Henceforth, gender marking across persons will be referred to as vertical marking.

Gender is not an exceedingly common grammatical category in the languages of the world. It has been estimated that only approximately one out of three languages possess sex-based nominal and/or pronominal gender (Berg 2020). However, if gender emerges, it may appear on almost all word classes, in all persons, and virtually in all numbers. It is observed on several structural levels ranging from the syntactic to the morphological and even the phonological. One of the major reasons why gender shows up on so many word classes is that it typically manifests itself as

529

an agreement phenomenon (e.g. Corbett 1991). The gender of the (unmarked) controller can be inferred from the overt gender marking on the target.

The focus on first-person singular gender restricts both the types of targets and the types of controllers to be subjected to scrutiny. Only those controllers are eligible which form a person paradigm. This criterion eliminates nouns, which are invariably third person, and leaves us with personal pronouns, possessive determiners, and possessive pronouns. The choice of targets is minimally confined to those that depend on controllers with a person paradigm. This requirement excludes articles and attributive adjectives, whose scope is within NPs. Pronouns do not take articles. This leaves us with predicative adjectives and verbs. Gender marking on the different word classes or different instantiations of the same word class will henceforth be referred to as horizontal marking.

Against this background, the present study sets out to probe possible inroads that the sex of the speaker may make into grammatical systems and to examine the proclivity of these components towards gender marking. The focus of this study on the sex of the speaker situates gender in the grammatical category of person, which is organized in paradigms. An analysis of first-person gender marking cannot therefore afford to ignore second- and third-person forms. This approach allows us to assess the consistency of paradigms. A consistent system is defined as exhibiting a gender contrast in all persons. Consistency is taken to be a descriptive principle, which itself is in need of an explanation. There are arguments for and against consistency. Paradigms may strive towards consistency because consistent systems would seem to be less complex and easier to learn than inconsistent ones. At the same time, consistent systems may involve more marking than inconsistent ones. Since marking requires more effort than no marking, inconsistent systems might be regarded as less complex than consistent ones.

First-person gender marking is at the crossroads of grammatical and indexical gender. Rose (2018) argues convincingly that the two types are categorically distinct. This can be most clearly seen in third-person pronouns, which may not only distinguish between male and female referents but also vary according to the sex of the speaker. In Kayabí (Tupi-Guaraní), for example, the masculine singular form 'he' is $\tilde{g}a$ in male speech but $k\tilde{\iota}\tilde{a}$ in female speech. The feminine singular 'she' is rendered as $\tilde{e}\tilde{e}$ by men and as *kyna* by women (Dobson 1997: 13). This quadruple could logically be observed in all persons of a paradigm but one. Because the first person conflates sex and gender owing to its self-reference, the maximum number of forms we may expect to find here is two.[2] For these two forms, it is difficult to distinguish between grammatical and indexical gender. Drawing on the notion of consistency, Rose (2018) suggests that first-person gender should be viewed as grammatical if the second and third person show a contrast in grammatical gender. If, however, the second- and third-person pronouns show a contrast in indexical gender, then first-person gender should be regarded as indexical.

---

[2] More complex systems in which first-person pronoun choice is determined by the sex of the speaker as well as that of the addressee also exist but are extremely uncommon (see Rose 2018).

While Rose's proposal makes good sense, uncertainties surrounding the proper classification of first-person gender remain. This is particularly true of those languages that do not mark gender in all persons. Since the chief descriptive aim of this paper is to assemble a maximum number of cases of first-person gender, it seems wise to deploy an inclusive sampling strategy and to stay agnostic about the theoretical interpretation of the first-person gender contrast.

This work is structured as follows. Section 2 deals with methodological issues, particularly with the twin methods of data collection. The empirical analysis in Section 3 focuses mainly on vertical consistency across the different hosts of gender marking. The theoretical discussion in Section 4 attempts to provide an explanation of the varying degrees of consistency that are found across word classes by relying on the distinction between referential and grammatical units. The conclusion in Section 5 highlights the ways in which first-person gender penetrates grammatical systems.

2. METHOD

In view of the assumed uncommonness of the first-person gender contrast, the net had to be cast wide. In fact, two complementary methods of data collection were employed, both of which are standard practice in typological research. They are complementary in the sense that one is more oriented towards breadth and the other more oriented towards depth. The first is the standard procedure of extracting information from relevant sources. Not only grammars and grammar sketches but also genealogical and areal overviews as well as more specific survey articles were perused (e.g. Forchheimer 1953, Ibriszimov & Segerer 2004). My overarching aim was to conduct as exhaustive a search as was possible within the limits of data availability. Therefore, I consulted all sources I could locate instead of creating a balanced sample by including certain languages and excluding others. It is more than likely that I missed a number of relevant languages, but it is equally likely that these languages constitute a minority and that their inclusion would not seriously affect the pattern of results. Overall, 1,750+ languages were checked.[3] This world-wide search yielded a total of 90 languages that mark first-person singular gender somewhere in their grammatical systems. These languages, along with the relevant references, are listed in the online materials.

The fact that reference grammars do not usually provide full person paradigms across all word forms (e.g. tense, aspect, and voice in verbs) called for a second method – data elicitation. A questionnaire was prepared in which native speakers were asked to translate a set of sentences into their mother tongue. This approach allowed for a more systematic collection of data as well as a comparison of functionally similar phenomena across different languages. It had to be decided beforehand which forms to include and which to exclude. This decision was based

---

[3] These are the languages that I checked myself. However, the surveys that I consulted do not always indicate how many languages were scrutinized.

on a perusal of the grammars of the languages for which I solicited translations. Clearly, I could only include those areas where I expected to find gender contrasts. Therefore, I have almost certainly missed out on a number of language-specific aspects that are sensitive to gender. Given the focus on paradigm consistency, all relevant word classes were examined for first-, second-, and third-person singular.

The questionnaire took the form of a blueprint that could, and had to be, adapted to the specifics of the individual language tested. In the simplest case, certain sentences had to be deleted because a grammatical category such as the passive just did not exist. Furthermore, certain words did not have a natural translation equivalent in the target language or involved a word class change. In discussions with the native speakers, these items were replaced with more suitable items. Such adaptations were also necessary if the original word was not susceptible to gender marking but a different member of the same word class was. The skeleton questionnaire can be found in the supplementary materials.

The questionnaire took the following variables into consideration: person (first vs. second vs. third), word class (verb vs. adjective), tense (present vs. past), valency (transitive vs. intransitive), and voice (active vs. passive). Of course, the inclusion of additional contrasts would have been desirable, but for purely practical reasons, the length of the questionnaire had to stay within limits.

All of the native speakers that I worked with on the questionnaire have a background in linguistics, which was essential for the segmentation into morphemes and the determination of their functions, i.e. the interlinear gloss. Some of my consultants are professional linguists.

The questionnaire was distributed to native speakers of 25 languages with nominal gender of which 21 belong to the Indo-European and four to the Afro-Asiatic stock, all of which are Semitic languages. The Indo-European languages classify into 17 European and four 'Indo' languages, which themselves divide into two Iranian and two Indic languages. The European languages branch into six Romance languages, seven Slavic languages, two Baltic languages, Albanian, and Greek. All these languages make up what is termed the 'questionnaire sample'.

This set of languages constitutes anything but a representative sample. Its composition is determined partly by availability of informants and partly by the fact that sex-based gender is very unevenly distributed across families and genera. The strong predominance of Indo-European languages in the sample ensues from these two factors. As a result of this bias, the questionnaire sample cannot claim to be a faithful reflection of gender coding in the languages of the world. It serves to provide a first idea of how and where first-person singular gender is preferentially coded. The evidence that it furnishes is suggestive rather than conclusive.

There is a certain division of labour between the two methods of data collection. Since a full account of the pronoun system is part and parcel of virtually all grammatical descriptions, the questionnaire did not deal specifically with pronouns. Instead, it focused on predicative adjectives and verb paradigms. However, whatever pertinent information could be gleaned from grammars was of course also taken into consideration.

The languages for which gender information was culled from the relevant literature and the languages of the questionnaire sample form non-overlapping sets.[4] Thus, the number of languages subjected to analysis amounts to 90 + 25 = 115 in total.

3. DATA ANALYSIS

The results of this study are presented across four thematic sections. Section 3.1 introduces personal pronouns, Section 3.2 possessive determiners, Section 3.3 predicative adjectives, and Section 3.4 verbs. Theoretical and methodological issues that are specific to the word class in question are taken up in the individual subsections.

3.1 *Independent personal pronouns*

This subsection focuses on independent personal pronouns to the exclusion of bound person markers. The former have the great advantage of occurring in all three persons in the vast majority of the world's languages. A thorough worldwide search for gender-specific first-person singular personal pronouns was conducted by consulting dedicated works such as Schmidt (1919a), Forchheimer (1953), and Ibriszimow & Segerer (2004) in addition to a mountain of reference grammars.

None of the pertinent literature is specifically devoted to first-person gender. Siewierska (2013) gives a typological profile of gender distinctions as a function of person and number. While she reserves a separate category for third-person pronouns, she combines first and second person. Neither Corbett (1991) nor Plank & Schellinger (1997) pay particular attention to first-person gender contrasts.

Table 1 lists all the languages with a gender distinction in first-person singular pronouns that I could find. It is enriched with information on the macroarea, the family, and the genus to which these languages appertain, taken from the World Atlas of Language Structures (WALS) for those languages which are included in this database (see Dryer & Haspelmath 2013) and from Glottolog 4.6 for the other languages.

A grand total of 30 languages invites a slightly ambivalent interpretation. On the one hand, this is a tiny fraction relative to the large number of languages that have been checked. Essentially, this is the result that would be predicted by the Redundancy Hypothesis referred to above. On the other hand, a total of 30 languages represent a much higher yield than what we find in Bhat's (2004: 109) and Siewierska's (2013) samples (both N = 0).[5] Clearly, these languages are too many to be written off as mere historical accidents. That is, the Redundancy Hypothesis alone is incapable of dealing with this finding. The first conclusion to be drawn from Table 1 is, then, that the obviousness of the sex of the speaker does not prevent the marking of first-person singular gender.

---

[4] Arabic and Tihamah Arabic (spoken in Yemen) were separately counted because they behave differently in one crucial respect.

[5] Siewierska's (2013) WALS chapter 44 does mention two languages with first-person gender, but these are not included in the sample on which her chapter is based.

	Language	Macroarea	Family	Genus	Source
1	Awetí	South America	Tupian	Awetí	de Cássia Borella 2000
2	Barupu	Papunesia	Skou	Warupu	Corris 2005
3	Burmese	Eurasia	Sino-Tibetan	Burmese-Lolo	Jenny & Tun 2016
4	Classical Tibetan	Eurasia	Sino-Tibetan	Bodic	Beyer 1993
5	Djingili	Australia	Mirndi	Djingili	Chadwick 1975
6	Ekari	Papunesia	Trans-New Guinea	Wissel Lakes-Kemandoga	Voorhoeve 1975
7	Fulniô (Yatê)	South America	Yatê	Yatê	da Costa 1999
8	Garifuna	South America	Arawakan	Caribbean Arawakan	de Pury 2003
9	Goajiro	South America	Arawakan	Caribbean Arawakan	Ramirez Gonzales 1995
10	Hadza	Africa	Hadza	Hadza	Sands 2013
11	Itonama	South America	Itonama	Itonama	Camp & Liccardi 1967
12	Japanese	Eurasia	Japanese	Japanese	Hinds 1986
13	Kala Lagaw Ya (Saibalgal)	Australia	Pama-Nyungan	Northern Pama-Nyungan	Schmidt 1919b
14	Karajá	South America	Macro-Ge	Karajá	Ribeiro 2012
15	Korana	Africa	Khoe-Kwadi	Khoe-Kwadi	Meinhoff 1930
16	Kukama-Kukamiria	South America	Tupian	Tupi-Guaraní	Vallejos 2016
17	Laal	Africa	Laal	Laal	Boyeldieu 1982
18	Máku	South America	Maku	Maku	Rivet et al. 1924-5
19	Ngala	Papunesia	Sepik	Ndu	Laycock 1965
20	Ofayé	South America	Macro-Ge	Ofayé	de Oliveira 2006
21	Omagua	South America	Tupian	Tupi-Guaraní	O'Hagan 2011
22	Páez	South America	Páezan	Páezan	Jung 1989
23	Rikbáktsa	South America	Macro-Ge	Rikbáktsa	Silva 2005
24	Shabo	Africa	Shabo	Shabo	Tsehay Taye 2015
25	Tainae	Papunesia	Trans-New Guinea	Angan	Carlson 1991
26	Ternate	Papunesia	North Halmaheran	North Halmaheran	Watuseke 1991
27	Thai	Eurasia	Tai-Kadai	Kam-Tai	Smyth 2002
28	Tihamah Arabic	Africa	Afro-Asiatic	Semitic	Behnstedt 1985
29	Tocharian A	Eurasia	Indo-European	Tocharian	Sieg & Siegling 1931
30	Tsafiki	South America	Barbacoan	Barbacoan	Moore 1979

Table 1
Languages with a gender contrast in first-person singular personal pronouns (N = 30).

A look at the family membership of the languages at hand reveals that first-person gender marking is not a one-off phenomenon. It is remarkable that there is no genealogical clustering in the data. The 30 languages are distributed across 23 different families (or constitute isolates). Three families (i.e. Arawakan, Sino-Tibetan, and Trans-New Guinea) are represented twice, and two families (i.e. Macro-Ge and Tupian) are represented three times. Moreover, with the exception of Garifuna and Goajiro as well as Kukama-Kukamiria and Omagua, no two languages belong to the same genus. This wide scatter of the languages suggests that first-person singular gender is a recurrent event that is motivated by the nature of individual languages and their sociocultural context and arose independently in many different places. In other words, it did not emerge once and proliferated within one and the same family or diffused across different families through language contact. This is a good reason to regard first-person gender as typologically significant.

The areal analysis confirms this conclusion. Of the six macroareas in WALS, five are attested in Table 1. While North America is not represented at all, South America is overrepresented. As many as 13 of the 30 languages are located in this macroarea (compare Rose 2015).[6] The other languages are distributed rather evenly across the remaining macroareas. Overall, first-person singular gender does not appear to be areally restricted.

Prior to examining the consistency of the personal-pronoun paradigms, it is instructive to study the formal relationship between the masculine and the feminine first-person variants. Three types of formal relationship are conceivable, with fuzzy boundaries between them, viz the suppletive, the phonological, and the morphological. These types will be illustrated on the basis of third-person pronouns. Two forms are classified as suppletive if they are formally dissimilar, as in I'saka (Skou) *kia* 'he' – *umu* 'she' (Donohue & San Roque 2004). A morphological relationship holds between two forms when at least one form is morphologically analyzable because it contains a discrete gender marker. This classification relies mainly on whether the original source provides a morphological analysis of the pronouns. The case is particularly clear when the assumed gender marker recurs in other members of the same paradigm or elsewhere. The morphological type may be subdivided into an additive and a substitutive type. The additive type may be illustrated by Slovak third-person pronouns in which the masculine form is suffixless (*on* 'he') while the feminine form is suffixed (*on-a* 'she'). The substitutive type may be exemplified by Sheko (Omotic) *ás-əra* 'he' – *íʃ-əra* 'she' (Hellenthal 2010).

A phonological relationship is defined by phonological similarity between monomorphemic masculine and feminine forms. Two items are categorized as phonologically similar if identical segments outnumber differing segments. An illustrative case comes from Murrinh-Patha (Southern Daly) *nukunu* 'he' - *niɣunu* 'she'

---

[6] This effect is no statistical artefact. South America contributes fewer languages to my sample (as well as to WALS and Glottolog) than would be expected by chance (1/6). Thus, correcting for chance even strengthens this bias.

(Mansfield 2019). Forms with equal numbers of identical and non-identical segments were also classed as phonologically similar because the baseline probability of finding identical segments is lower than that of finding non-identical segments.

The three types of formal relationship between the masculine and the feminine form make different predictions as to paradigm consistency. A morphological marker separates gender out from the pronominal base. This relative independence would allow the same (or even a different) gender marker to attach to other persons. Hence, the morphological type may be predicted to give rise to consistent paradigms. By contrast, the phonological and the suppletive types do not bring about such an independence. They are therefore predicted to give rise to less consistent paradigms.

In an attempt to test the Identity Hypothesis introduced in the opening section, the original sources were also searched for information on the connotations of first-person pronouns. It may very well be that this information is not provided even though there may be pragmatic overtones. It is possible, therefore, that in the following analysis, the true extent of these pragmatic cues is underrepresented. In several languages including Ekari and Fulniô, first-person sex-specific forms co-occur with first-person sex-neutral forms.

Table 2 presents the first-person singular forms of the languages listed in Table 1. Furthermore, it includes information on the type of formal contrast between the masculine and the feminine variant (morphological, phonological, suppletive), a possible gender contrast in the second and third person as well as additional meanings of the first-person pronoun (if available).

Prior to the analysis proper, it should be pointed out that Itonama diverges slightly from the other languages in having a common gender rather than a masculine-only form (i.e. *os-ni*). Yet, it is included in the database because gender does play a role in the first-person singular by distinguishing between common and feminine gender.

The 30 languages in Table 2 divide into 11 languages of the suppletive type, 10 languages of the phonological type, and nine languages of the morphological type. The suppletive type is exemplified by Tsafiki *la – chi'qué*, the phonological type by Karajá *diar̃ – dikar̃*, the additive morphological type by Hadza *'ono – 'ono-ko*, and the substitutive morphological type by Korana *ti-re – ti-ra*, where *–re* (or just *–r*) is the masculine marker and *–ta* the feminine marker. The additive and the substitutive types are about equally frequent (five vs. four languages).

The rather even distribution of the three types of formal relationship is noteworthy. It suggests that each type has its own 'supporters' and that this support is of similar strength. Suppletion is known to be highly correlated with frequency. Since personal pronouns are high-frequency words, they are highly likely to be suppletive not only across persons but also within persons. The phonological similarity

---

[7] The gender contrast in Garifuna appears to be on its way out of the language. Haurholm-Larsen (2016) notes the complete absence of independent personal pronouns in present-day Garifuna. Of course, this leaves no room for a gender distinction (see Dunn 2015 on the diachronic instability of genderlects and Güldemann 2004 on the emergence of first-person gender).

[8] Tidore, another North Halmaheran language, has exactly the same pair of first-person gendered pronouns (van Staden 2000). It was therefore not included in the database.

	Language	Male Speech	Female Speech	Pragmatics	Type	Contrast in 2 SG	Contrast in 3 SG
1	Aweti	atit	ito		sup	no	no
2	Barupu	nená	ненí		pho	yes	yes
3	Burmese	tɕənɔ	tɕəmá	polite	pho	yes	yes
4	Classical Tibetan	kho-bo	kho-mo		mor	no	yes
5	Djingili	ŋaini	ŋaigja		pho	yes	yes
6	Ekari	ani ki	ani ko	non-neutral	mor	yes	yes
7	Fulniô (Yatê)	la	o:so, owe-so		mor	yes	yes
8	Garifuna[7]	aú	nugaya		sup	yes	yes
9	Goajiro	taja-kai	taja-kat		mor	yes	yes
10	Hadza	'ono	'ono-ko		mor	yes	yes
11	Itonama	os-ni	os-ni-ʔka		mor	yes	yes
12	Japanese	boku	atashi		sup	yes	yes
13	Kala Lagaw Ya (Saibalgal)	nai	nazo		pho	no	yes
14	Karajá	diarɔ̃	dikarɔ̃		pho	no	no
15	Korana	ti-re	ti-ta		mor	yes	yes
16	Kukama-Kukamiria	ta	etse, tsa		pho	no	yes
17	Laal	já	jí		pho	no	yes
18	Máku	ax	ax̩-e̩i		mor	yes	yes
19	Ngala	wn	ñən		sup	yes	yes
20	Ofayé	a	ag^n		pho	yes	yes
21	Omagua	tá	tsí		sup	no	yes
22	Páez	ad	u'kwe		sup	yes	–
23	Rikbáktsa	u'ta	(i)ki'ra		sup	no	yes
24	Shabo	tiŋ	ta		sup	yes	yes
25	Tainae	to	te		pho	yes	yes
26	Ternate[8]	fangare	fajaru	polite	sup	no	yes
27	Thai	phŏm	(di)chán	polite	pho	no	no
28	Tihamah Arabic	ʔan-a	ʔan-i		mor	yes	yes
29	Tocharian A	näṣ (nṣ-äk)	ñuk		sup	no	?
30	Tsafiki	owe	chi'qué		sup	no	no

Table 2
First-person singular personal pronouns in male and female speech.

between masculine and feminine forms can be taken to iconically reflect their discourse-functional similarity. The morphological type may also be iconically motivated. It may evidence a one-to-one correspondence between form and meaning in that one morpheme denotes the first person and the other the sex of the speaker.

We move on to consider the person paradigm. Of the 30 languages, 18 show a gender contrast in the second person, while 12 fail to do so. Two languages had to be discarded from the third-person count. The third-person pronoun is zero-marked in Páez and unknown in Tocharian A. Of the remaining 28 languages, 24 exhibit a grammatical-gender contrast while only four do not.[9] As many as 17 of the 28 languages evince a gender contrast in both second and third person.

It is not surprising to observe a higher rate of gender marking in third than in second person. This is also true of the vast majority of languages that do not code gender in the first person. That is, the languages in Table 2 are highly regular in this respect. However, the rate of gender marking in second and third person is much higher in these languages than in those with a gender-neutral first-person pronoun. On the basis of Siewierska's (2013) sample, the rate of gender marking may be estimated to be 32% for third-person singular and 5% for second-person singular. These figures contrast sharply with a rate of 86% for third person and 60% for second person in Table 2. The conclusion invited by these data is that gender-specific first-person pronouns are integrated into a paradigm with an elevated rate of gender marking in the other persons. The occurrence of first-person gender appears to be facilitated by second- and third-person gender. Obviously, this elevated rate of gender marking leads to a heightened consistency in the paradigms of the languages listed in Table 2. It should be borne in mind that this is a probabilistic rather than a categorical effect.

The preceding analysis probed into the relationship between the masculine and the feminine form as well as that between gender and person. It is now time to pull the two strands together and enquire into a possible interaction of the type of formal contrast and the extent of gender marking in the person paradigm. The pertinent results are summarized in Table 3. For reasons explained above, Páez and Tocharian A had to be excluded from this analysis.

While the numbers are too low for a definitive statement, two claims are compatible with the data. The phonological and the suppletive types behave similarly in showing only a weak trend towards paradigm consistency. In contrast, the morphological type favours a gender contrast in all persons. That is, consistency is greatly facilitated by morphological gender marking. This is as predicted. The relative independence of a (bound) gender morpheme encourages its recurrent use across persons, whereby gender consistency is established. It seems, then, that there is an effect of the type of formal relationship on gender consistency. The morphological type stands out in that it creates paradigm consistency more successfully than the other two types do.

---

[9] Aweti and Karajá mark indexical gender, though not grammatical gender in the third person singular. In an effort to avoid the mixing of grammatical and indexical gender, these two languages were coded as possessing no (grammatical) gender in the far-right column of Table 2.

538

Extent of Gender Marking	Both Second and Third Person	Only Third Person	Neither Second nor Third Person	Consistency
Type of formal contrast in 1 SG				
Morphological	8	1	0	8:1
Phonological	5	4	1	5:5
Suppletive	4	2	3	4:5

Table 3
Number of languages by type of formal contrast and extent of gender marking (N = 28).

Table 2 provides some evidence for a connection between gender marking and pragmatic function. In four of the 30 languages, the relevant grammar includes a remark to the effect that the gender-specific pronoun is part of a paradigm with special pragmatic overtones where politeness plays a pre-eminent role (e.g. 'I, as a polite speaker'; Bhat 2004: 112). While politeness is generally more listener- than speaker-oriented, it may also be linked with the sex of the speaker in that for a male speaker, politeness may not be the same as politeness for a female speaker (e.g. 'I, as a polite female speaker'). It may be suggested that a pragmatically marked form is more likely to code gender than a pragmatically unmarked one is.

The few languages with pragmatically motivated first-person gender indicate that the Identity Hypothesis can only account for a small part of the data. It seems likely, therefore, that the redundancy introduced by first-person gender is facilitated by other factors such as paradigm consistency and its processing benefit.

Finally, we will briefly consider how two prominent person hierarchies fare in light of the above data. Siewierska (2004) propounds the person hierarchy 3 > 2 > 1, which allows for three language types: (i) languages with gender in all three persons, (ii) languages with gender in the second and third person, and (iii) languages with gender in the third person only. Siewierska herself notes that not all languages fall into one of these patterns. Her implicational hierarchy is a stronger version than Greenberg's (1963) Universal 44, which states that a gender contrast in the first person presupposes a gender contrast in the second and/or third person.

Strictly speaking, the data in Table 2 conform neither to Siewierska's hierarchy nor to Greenberg's Universal. There are two languages, namely Tsafiki and Thai, which have neither second- nor third-person gender. And there are as many as 10 further languages which lack second-person gender. Thus, a gender contrast in the first person does not presuppose a gender contrast in the second. This conclusion is not compatible with a rigorous interpretation of Siewierska's hierarchy.

Greenberg's Universal does not fare much better. The crux is that it blurs the distinction between second and third person. However, this distinction should be upheld because second- and third-person gender relate to first-person gender in different ways and vary in the extent to which they contribute to paradigm consistency. Whereas only 60% of the languages in Table 2 possess a gender contrast in the second person, the great majority of languages do so in the third

539

person.[10] Thus, there are exceptions to two of the three constellations envisaged by Greenberg. In Table 2, we find languages lacking second-person gender and languages lacking both second- and third-person gender.

3.2 *Possessive determiners*

While both possessive determiners and possessive pronouns form a person paradigm, only determiners will be considered here. This is because many grammars do not explicitly distinguish between the two functions or do not mention genuine possessive pronouns at all. Owing to the relational nature of possession, possessive determiners may host two types of gender – the gender of the possessor and that of the possessum. Naturally, we are only concerned with the former type. In first-person singular possessive determiners, the possessor is the speaker, so possessor gender in the first person is indicative of the sex of the speaker.

The focus of the present analysis on grammatical systems implies that only those languages are eligible which mark the gender of the possessor throughout the lexicon, i.e. independently of individual lexical items. In many languages, possessive determiners are formed on the basis of personal pronouns. The four processes by which the former are typically derived from the latter are conversion, affixing, the addition of a possessive particle, and the selection of a member from the case paradigm of personal pronouns (most usually the genitive). Which morphological process is preferentially chosen may give us a clue as to how the gender feature comes into being in those languages that have gender-specific personal pronouns. Logically speaking, there are two options. Gender may either develop anew in possessive determiners or be copied from the personal pronouns. A look at the morphological relationship between the two sets will allow one to determine which of the two options is taken. If this relationship is a suppletive one, the evidence will weigh in favour of gender developing anew in this domain. If, by contrast, personal pronouns are converted into possessives, the gender feature will be copied along the way. Basically, the same reasoning applies to affixing. If a possessive affix (or a free-standing particle) is added to create possessive determiners, the gender distinction is bought wholesale from the personal pronouns.

As in the analysis of personal pronouns, the search for gender-specific possessive determiners relied entirely on published (and unpublished) sources. Table 4 lists all the languages I could find with a possessor-gender contrast in first-person singular forms. In addition to the masculine and feminine forms, it includes the type of

---

[10] Rose (2013) made an attempt to rescue both Greenberg's and Siewierska's hypotheses by assigning them to partly different domains. Greenberg's Universal is claimed to hold when applied to grammatical gender alone while Siewierska's hierarchy is claimed to apply to grammatical and indexical gender alike. The problem with this assumed division of labour is two-fold. For one thing, it still leaves us with a few exceptions, as Rose herself notes; for another, it does not really rescue Universal 44, which can be subsumed under Siewierska's hierarchy because the domain of the former is a proper subset of the domain of the latter.

540

	Language	Male Speech	Female Speech	Morphological Process	Contrast in Second Person	Contrast in Third Person
1	Barupu	nená	není	conversion	yes	yes
2	Burmese	tɕənɔ́	tɕəmá	conversion (+ creakiness)	yes	yes
3	Hadza	'one-ne-ya[11]	'one-te-ya	suffixing	yes	yes
4	Inanwatan	náre-so	naridó-wo	-	yes	yes
5	Japanese	boku no	atashi no	possessive particle	yes	yes
6	Kala Lagaw Ya (Saibalgal)	nau	uzu	phonological change	no	yes
7	Korana	ti-re	ti-ta	conversion	yes	yes
8	Kukama-Kukamiria	ta=	ca=	conversion	no	yes
9	Páez	ad	u'kwe	conversion	yes	–
10	Ternate	fangare ri	fajaru ri	possessive particle	no	yes
11	Thai	phǒm	(di)chán	conversion	no	no
12	Tocharian A	ñi	nañi	genitive	no	?
13	Tsafiki	la-chi	chi'qué-chi	suffixing	no	no
14	Yaqay	angga^{e}pape	angga^{u}pape	-	yes	yes

Table 4
First-person singular possessive determiners in male and female speech (N = 14).

morphological process (if applicable) as well as an indication of whether a gender contrast occurs in second and third person.

A comparison between Tables 2 and 4 reveals a striking similarity: most of the languages in Table 4 also figure in Table 2. That is to say, languages with a gender contrast in possessives are highly likely to also have a gender contrast in personal pronouns. This is true of 12 of the 14 languages listed in Table 4. This effect may be formulated as a weak implicational universal: if a language distinguishes gender in the first person of possessive determiners, it is likely to do so in personal pronouns. Differently put, a language without a gender contrast in first-person personal pronouns is unlikely to have it in first-person possessives.

The basis for this implicational universal lies in the morphological relationship between the personal pronouns and the possessive determiners. As Table 4 shows, suppletion (as in German *ich* [iç] 'I' vs. *mein* [maɪn] 'my') is completely absent. The possessive determiners in almost all relevant languages arise through conversion or through the addition of a possessive marker to the personal pronouns. Hence, the languages in Table 4 evince a gender contrast because they get gender 'for free'. Viewed from the opposite angle, possessive determiners would seem unlikely to

---

[11] These are the forms appropriate for masculine possessums. The forms for feminine possessums are structurally quite similar.

541

develop a gender contrast through processes of grammaticalization that are unique to this domain. This underscores the fact that gender marking in this domain is rare in the absence of gender marking on first-person personal pronouns.

The comparison between Tables 2 and 4 further reveals that Table 4 contains less than half as many languages as does Table 2 (even if the 'newcomer' languages are added in). This difference suggests that a language with a gender distinction in first-person personal pronouns stands a less than 50% chance of preserving this distinction in first-person possessives. Thus, possessive determiners are inherently less likely to code first-person gender than personal pronouns are. On the logic that the possessive determiners are derived from the personal pronouns, this may be taken to argue for a certain vulnerability of gender.

The two languages in Table 4 which lack a gender distinction in first-person personal pronouns are Inanwatan (South Bird's Head) and Yaqay (Marind). In Inanwatan, all three persons are expanded by the gender suffixes –so (masculine) and –wo (feminine) (de Vries 2004: 29). This is a case of paradigm consistency as a result of morphological gender marking (see previous subsection). A similar analysis applies to Yaqay (Boelaars 1950: 61).

The two rightmost columns in Table 4 permit us to study gender marking in the second and third person, respectively. As before, Páez and Tocharian A are left out of account. The remaining 12 languages show a clear predilection for gender marking, which is somewhat stronger in the third than the second person. Gender is marked in the third person by 10 languages and in the second person by eight languages. That is, more than half of the languages (seven out of 12) exhibit consistent gender coding. It may be argued, therefore, that the occurrence of first-person gender should be viewed against the background of an elevated sensitivity to gender marking in the possessive paradigms of the languages at hand.

Concluding, languages with a gender distinction in first-person possessive determiners are extremely uncommon – more uncommon, in fact, than languages with a gender distinction in first-person personal pronouns. If a gender contrast is found in possessive determiners, it is usually inherited from personal pronouns. This inheritance is guaranteed by the process of conversion in which, by definition, nothing is gained and, more importantly, nothing is lost. The other facilitating factor is the paradigm. First-person gender finds a more natural home in paradigms with an above-average sensitivity to gender marking.

3.3 Predicative-adjective constructions

The obvious reason that predicative adjectives are included in this study is that they may be gender-marked. At the same time, they are not marked for person. Therefore, they cannot be argued to form a person paradigm. However, they are controlled by the pronominal subject which does form a paradigm. This grammatical connection raises the possibility of an effect of the person paradigm on gender marking on the adjective. On the one hand, the lack of person indices leads us to

542

expect gender marking across the board. On the other hand, the Redundancy Hypothesis predicts inconsistent marking.

It will be seen that the loci of gender marking in predicative-adjective constructions vary across languages. Gender coding may occur not only on the inflecting adjective itself but also on other word classes. In fact, four different types of targets of first-person gender marking could be identified. A separate subsection will be devoted to each of them. Owing to the varying nature of these gender sites, the predictions regarding gender coding will be formulated afresh in each case. There is no claim that the four types of gender-marking units in predicative-adjective constructions are the only possible ones.

From now on, a great deal of reliance will be placed on the questionnaire data. These data are occasionally supplemented with data from the relevant literature.

3.3.1 *Gender on predicative adjectives*

Gender marking on adjectives is quite common in gendered languages. The following example is from Latvian, a Baltic language.

(1) Latvian: a. Viņš ir laimīg-s.
   3SG.M BE.3SG happy-M
   'He is happy.'
   b. Viņa ir laimīg-a.
   3SG.F BE.3SG happy-F
   'She is happy.'

In addition to the gender distinction between the third-person singular pronouns *viņš* 'he' and *viņa* 'she', gender marking occurs on the inflected adjective: the suffix –*s* flags masculine gender, while –*a* flags feminine gender.

As noted above, the absence of person indices on the adjective in (1) may be assumed to imply person neutrality and hence to produce consistent gender marking. However, this is not altogether clear at the outset. Gender marking on personal pronouns tends to disappear as we shift from the third to the first person (see Section 3.1). It may be that the same reason that militates against gender marking on first-person pronouns (the Redundancy Hypothesis) reduces the likelihood of gender marking on adjectives in first-person subject sentences. Thus, we face two conflicting predictions. The absence of person indices on adjectives predicts consistent marking, whereas the principle of redundancy might give rise to inconsistent marking. Let us see which route Latvian has chosen.

(2) Latvian: a. Es esum laimīg-s.
   1SG BE.1SG happy-M
   ♂: 'I am happy.'
   b. Es esum laimīg-a.
   1SG BE.1SG happy-F
   ♀: 'I am happy.'

543

As can be seen, the gender distinction on the adjective in the third person in (1) is replicated in the first person in (2). The same contrast emerges in the second person. Thus, consistency prevails.

The languages in the questionnaire sample are strikingly unanimous. All languages that evidence a gender contrast in the third person (N = 25) also evidence a gender contrast in the first and second person in predicative-adjective constructions. We thus observe widespread consistency.

There are more languages out there with gender-specific adjectives, both within Indo-European and beyond. For example, adjectives in several Indic languages inflect for gender. One such language is Marathi in which we find the contrast between *moth-ā* 'big-M' and *moth-i* 'big-F' (Pandharipande 1997: 451). In personal communication, the author points out that this variation occurs in all persons, thereby confirming the result obtained for the questionnaire languages. The same is true of Coastal Marind, a language from the Anim family, in which the masculine form of 'small' is *papes* and the feminine form *papus* in all three persons (Bruno Olsson, p.c.). In point of fact, I did not come across a grammar that proclaims gender inconsistency. However, the lack of documentation of full-person paradigms cannot automatically be taken to imply that all persons behave alike.

To conclude, even though the questionnaire languages cannot speak for the languages of the world, there is good evidence that consistency is a characteristic feature of gender marking on predicative adjectives. It is notable that the presence or absence of gender marking on personal pronouns plays no role in this decision. Thus, gender specifications on predicative adjectives appear to be independent of gender marking on the subject.

3.3.2 *Gender on the indefinite article*

In predicative-adjective constructions, the subject and the adjective are often linked up by a copular verb. I found one language, namely Beja (Cushitic), in which the copula is augmented by an indefinite article that carries gender information (Vanhove 2017). As indefinite articles are not marked for person, they are predicted to code gender throughout. Wedekind, Wedekind & Musa's (2005) pedagogical grammar of Beja contains full-person paradigms. Example (3) is borrowed from Wedekind et al. (2005: 55).

(3) Beja: a. Ani eeraa-b-u.
            1SG white-INDEFINITE.M-BE.1/3SG[12]
            ♂: 'I am white.'

---

[12] Wedekind et al. do not consider –*u* marked for person. However, the fact that the second-person suffixes differ from the first- and third-person suffixes suggests that person information is not completely absent from the suffixes. This analysis agrees with Vanhove's (2017) grammatical description.

    b. Ani eeraa-t-u.
        1SG white-INDEFINITE.F-BE.1/3SG
        ♀: 'I am white.'

As can be gathered from (3), the indefinite article on the adjective indexes the sex of the speaker. The /b/ in (3a) identifies a male speaker, while the /t/ in (3b) identifies a female speaker. This gender contrast pervades the person paradigm for both singular and plural forms in Beja predicative adjectives. In line with the prediction, gender is marked independently of person.

3.3.3 *Gender on the definite article*

Albanian is the only language in the questionnaire sample to mark gender (in all persons) on definite articles that obligatorily accompany predicative adjectives. Note, in passing, that in their adnominal use, definite articles are suffixed to their head nouns and that the phonological form of the adjectival and nominal use is only partly identical. Since definite articles are not marked for person, it is highly likely that gender is consistently marked. Refer to (4).

(4) Albanian: a. Unë jam   i   trishtwar.
                 1SG   BE.1SG ART.M sad
                 ♂: 'I am sad.'
             b. Unë jam   e   trishtwar.
                 1SG   BE.1SG ART.F sad
                 ♀: 'I am sad.'

The sex of the speaker is expressed by the choice of the definite article, which precedes the predicative adjective. As in the nominal use type, the masculine article is /i/ (see (4a)). Unlike the nominal use type, which takes /a/, the feminine article is /e/ (see (4b)). This strategy of gender marking is encountered throughout the person paradigm.

    For the sake of completeness, it may be added that not all Albanian adjectives follow this pattern. Another class consists of adjectives such as *budalla* 'stupid' and *simpatik* 'likeable' which mark gender by means of a suffix, just as in Latvian (see (1) and (2) above). The important point in the present context is that this morphological strategy is also consistently employed across persons.

3.3.4 *Gender on ezafe*

Kurmanji Kurdish is the only language in the questionnaire sample to mark gender by means of ezafe, a kind of connective found in a number of Iranian languages. There are different types of ezafe – in particular, independent particles and bound morphemes (see Haig 2019). As in Beja (see Section 3.3.2), the copula is suffixed to the adjective. Note that Kurdish lacks definite articles; the unmarked forms of nouns are inherently definite.

The prediction regarding first-person gender marking is obvious: since ezafe does not inflect for person, inconsistent marking is not to be expected. Anything but first-person gender marking would come as a surprise. Consider (5).

(5) Kurdish: a. Ez   yê   keyfxos-im.
                  1SG   EZAFE.M   happy-BE.1SG
                  ♂: 'I am happy.'
             b. Ez   ya   keyfxos-im.
                  1SG   EZAFE.F   happy-BE.1SG
                  ♀: 'I am happy.'

As shown in (5), the sex of the speaker is indicated by *yê* for males and *ya* for females. This pattern is replicated in the two other persons. As all personal pronouns are gender-neutral, ezafe is the only site where gender is marked. As predicted, Kurmanji Kurdish codes gender consistently.

3.3.5 *Interim summary*

All of the languages in the questionnaire sample code gender in predicative-adjective constructions. Whether this is a general characteristic of gendered languages worldwide remains to be seen. The four constructional types in which gender marking occurs are all highly consistent: whether gender marking materializes as a suffix, on an indefinite article, on a definite article, or on ezafe as an independent unit, it is observed throughout. Consistency is affected neither by whether the gender marker is bound or free nor by whether the personal pronoun is gender-specific or gender-neutral. Note that this is not necessarily so. Dropping the gender marker in the case of (pragmatic) redundancy is certainly a theoretical option.

3.4 *Verbs*

In a great many languages, verbs boast a larger number of inflectional variants than any other word class does. This plethora of forms allows us to study gender marking in greater detail. Logically, each tense, aspect, mood, or voice value of a verb may take an independent decision for or against gender marking. Unlike adjectives, verbs may carry person indices. Thus, they provide an opportunity of investigating both vertical and horizontal gender marking.

The following analysis will be confined to verb inflections. When the pronominal subject is compulsory, these are cases of subject-verb agreement. When the pronominal subject is optional, these inflections are more adequately referred to as person indices (Haspelmath 2013). Note that gender may also be coded on independent words in VPs. Such is the case in progressive markers in Shekhawati (Indic; Gusain, p.c.). Since these gender markers do not occur on inflected verbs, they do not stand a fair chance of interacting with person. Therefore, they were ignored.

546

3.4.1 *Vertical gender marking in the questionnaire sample*

The 25 questionnaire languages fall into three groups: (i) languages that consistently mark gender on verbs in all three persons, (ii) languages that do not mark gender at all, and (iii) languages that mark gender in the second and third person, though not in the first. This is a surprising set of results in that there is no language in the sample that marks gender in the third person only. Here is an example from the first group.

(6) Polish: a. (Ja)   gra-ł-e-m   na   pianinie.
                (1SG)   play-PAST-M-1SG on   piano
                ♂: 'I played the piano.'
          b. (Ja)   gra-ł-a-m   na   pianinie.
                (1SG)   play-PAST-F-1SG on   piano
                ♀: 'I played the piano.'
          c. (Ty)   gra-ł-e-ś   na   pianinie.
                (2SG)   play-PAST-M-2SG on   piano
                'You played the piano.' (male addressee)
          d. (Ty)   gra-ł-a-ś   na   pianinie.
                (2SG)   play-PAST-F-2SG on   piano
                'You played the piano.' (female addressee)
          e. (On)   gra-ł-ø   na   pianinie.
                (3SG.M) play-PAST-3SG.M on   piano
                'He played the piano.'
          f. (On-a)   gra-ł-a   na   pianinie.
                (3SG-F) play-PAST-3SG.F on   piano
                'She played the piano.'

As can be seen in (6), the Polish verb *grać* 'to play' distinguishes gender in all three persons in the past tense. Gender marking on the verb is independent of gender marking on the (optional) pronoun. Vertical consistency in gender marking predominates in all Indo-European languages in the questionnaire sample.

Inconsistent gender marking on the verb is found in Western Aramaic, as illustrated in (7).

(7) Western Aramaic: a. Ono   mshtareli   piano.
                        1SG   play.PAST   piano
                        'I played the piano.'
                    b. Hat   mshtarel<u>o</u>kh   piano.
                        2SG   play.PAST.M   piano
                        'You played the piano.' (male addressee)
                    c. Hat   mshtarel<u>a</u>kh   piano.
                        2SG   play.PAST.F   piano
                        'You played the piano.' (female addressee)

    d. Hie   mshtar'ele   piano.
       3SG.M   play.PAST.M   piano
       'He played the piano.'
    e. Hia   mshtar'ela   piano.
       3SG.F   play.PAST.F   piano
       'She played the piano.'

Example (7a) attests to the absence of gender marking in the first person. By contrast, the second person distinguishes between a male (7b) and a female (7c) addressee by inserting different vowels in the verbal template. Gender on the verb is also marked in the third person (see (7d) and (7e)). As in Polish, gender marking on the verb is independent of gender marking on the pronoun. Other Semitic languages such as Amharic and Arabic mark gender on second-person verb forms and second-person pronouns, while Western Aramaic does so only on verbs (see (7b) and (7c)). It seems that the status of gender marking on second-person verb forms differs from that of gender marking on second-person pronouns.

    The fact that not a single language in the sample has a gender contrast in the first person but none in the second person is certainly what the person hierarchy would lead us to expect. The occurrence of languages like Western Aramaic, for example, with a gender contrast in the second and third person, but none in the first person, is also entirely expected.

    Our interim conclusion is that vertical gender marking on verbs is characterized by a pronounced tendency towards paradigm consistency. The only case of inconsistency comes from the Semitic languages, which are consistent between the second and third person but inconsistent between the first and the second. This suggests that the relationship between first- and second-person gender marking may not be the same as that between second- and third-person gender marking. An additional observation pertaining to paradigm consistency comes from the personal pronouns. Whether these subject pronouns are gender-specific or gender-neutral does not seem to affect gender marking on verbs.

3.4.2 *Horizontal gender marking in the questionnaire sample*

We turn to a brief analysis of the various verb forms. Valency had no effect on gender marking. There was a minor effect of tense in that past tenses were more likely to code gender than present tenses. This effect is observed in some Slavic languages, as illustrated in (8) from Ukrainian.

(8) Ukrainian: a. Ya   hra-ju   na pianino.
               1SG play-PRESENT.1SG on piano
               'I play the piano.'
          b. Ya   hra-v   na pianino.
               1SG play-PAST.1SG.M on piano
               ♂: 'I played the piano.'

548

    c. Ya   hra-la   na pianino.
        1SG   play-PAST.1SG.F   on piano
        ♀: 'I played the piano.'

While no distinction between masculine and feminine forms is made in the present tense (8a), gender emerges in past tense forms. Masculine gender is expressed by –*v* in (8b), whereas feminine gender is expressed by –*la* in (8c). The same situation is obtained in Russian and Polish. There is no language in the sample in which the present tense is gender-marked while the past tense is not.

The major factor influencing gender marking, which cuts across tense and voice, is the distinction between simple and composite forms. The former are inflectional variants of a main verb, whereas the latter are created by combining an auxiliary with a participle. While simple tenses show little inclination towards gender marking, composite tenses are highly likely to mark gender. In fact, all composite tenses are gender-marked in the questionnaire sample. Gender marking always occurs on the participle but may also occur on both participle and auxiliary. It transpires that gender marking is preferentially found in composite forms but is not restricted to these.

It is noteworthy that the issue of gender marking seems to be entirely determined by form – not by function. Let us take the case of the passive, which is frequently gender-marked in the sample. However, gender marking is not brought about by the passive itself but by the fact that the passive is often a composite form. When it is a simple form, as in Ukrainian, no gender marking is observed. The same goes for tense. If the tensed verb is a simple form, it is unlikely to carry gender information (but see (8) above); if, however, it is a composite form, it is highly likely to do so. Consider the following examples.

(9) Spanish: a. (Yo)   golpe-o un coche.
               (1SG)   hit-1SG a   car
               'I am hitting a car.'
           b. (Yo)   fui   golpead-o por un coche.
               (1SG)   BE.PAST.1SG hit.PRTC-M by   a   car
               ♂: 'I was hit by a car.'
           c. (Yo)   fui   golpead-a por un coche.
               (1SG)   BE.PAST.1SG hit.PRTC-F by   a   car
               ♀: 'I was hit by a car.'

(10) Urdu: a. Main   piano   khelt-a   hoon.
              1SG   piano   play.PRTC-M BE.1SG
             ♂: 'I play the piano.'
          b. Main   piano   khelt-i   hoon.
              1SG   piano   play.PRTC-F BE.1SG
             ♀: 'I play the piano.'

549

Spanish is a good representative of the Romance languages in which the passive is composed of an auxiliary and the past participle of the main verb (see (9)). The Urdu case in (10) exemplifies gender marking in the present tense, which may now be argued to follow from the fact that the present tense is a composite form. In both (9) and (10), gender marking occurs by a dedicated suffix on the participle. In all cases in which first-person gender is marked on the participle, there is a paradigm consistency effect.

3.4.3 *Vertical gender marking beyond the questionnaire sample*

The final analysis leads us to consider languages other than those included in the questionnaire sample. Its major objective is to check whether the results of the questionnaire sample generalize to other languages – in particular, to those of non-Indo-European lineage. The focus is on vertical consistency.

A worldwide search yielded 56 additional languages with first-person singular gender in verb inflection. Languages for which grammars do not explicitly note or exemplify first-person gender – even though such can be inferred from the pertinent passages – were left out of account. Table 5 lists only non-Indo-European languages that code gender on verbs.[13] It includes only one representative of each genus.[14] Four of the languages in Table 5 (i.e. Barupu, Djingili, Goajiro, and Korana) already appeared in Table 1.

The languages in Table 5 are rather diverse. They come from all six macroareas and belong to 21 different families and 26 different genera. This is further evidence in favour of the claim that first-person gender marking is in no way areally restricted.

Gender marking on the verb is illustrated with examples from Tayap (Kulick & Terrill 2019: 85) in (11) and Coastal Marind (Olsson p.c.) in (12).

(11) Tayap: a. ŋa mbo-t b. ŋa mbo-k
                  1SG went-M   1SG went-F
                  ♂: 'I went.'   ♀: 'I went.'

---

[13] It therefore excludes the following Indo-European languages: Indic: Bhadrawahi (Dwivedi 2013), Bhalesi (Varma 1948), Gujarati (Doctor 2004), Hadothi (Dwivedi 2012), Kashmiri (Wali & Koul 1997), Konkani (Miranda 2007), Lamani (Barikeri 1982), Marathi (Pandharipande 1997), Palula (Liljegren, 2016), Pashayi (Lehr 2014), Shekhawati (Gusain 2001), Shumashti (Morgenstierne 1945), and Sirajic (Parihar & Dwivedi 2019). Iranian: Waneci (Elfenbein 1972). Romance: Ripatransone (Harder 1988).

[14] The following languages were additionally excluded from Table 5: Akhvakh (Avar-Andic-Tsezic; Creissels 2008), Angaataha (Angan; Lloyd 1973), Archi (Lezgic; Bond & Chumakina 2016), Assyrian Neo-Aramaic (Semitic; Khan 2008a), Dargwa (Lak-Dargwa; Forker 2020), Godoberi (Avar-Andic-Tsezic; Kibrik 1996), Guanano (Tucanoan; Stenzel 2013), Hamtai (Angan; Oates & Oates 1968), Hulaulá (Semitic; Khan 2004), Khwarshi (Avar-Andic-Tsezic; Khalilova 2009), Koy Sanjaq (Semitic; Mutzafi 2004), Kubachi (Lak-Dargwa; Vamling & Tchantouria 1991), Lishán Didán (Semitic; Khan 2008b), Lishanid Noshan (Semitic; Khan 1999), Siriano (Tucanoan; Criswell & Brandrup 2000), Tucano (Tucanoan; West 1980), Tuyuca (Tucanoan; Vlcek 2016), Wa'ikhana (Tucanoan; Waltz & Waltz 2015), and Yuruti (Tucanoan; Kinch & Kinch 2000).

    c. yu   mbo-t   d. yu   mbo-k
       2SG   went-M   2SG   went-F
      'You went.' (male addressee) 'You went.' (female addressee)

    e. ŋɨ   mbo-t   f. ŋgu   wo-k
       3SG.M   went-M   3SG.F   went-F
      'He went.'   'She went.'

(12)   Coastal Marind: a. e-pa-no-ihanot-a.
                     M-ABSC-1SG.ACTOR-run.away<1.UNDERGOER-EXTENDED
                     ♂: 'I am running (here where you are not looking).'
               b. u-pa-no-ihanot-a.
                     F-ABSC-1SG.ACTOR-run.away<1.UNDERGOER-EXTENDED
                     ♀: 'I am running (here where you are not looking).'
               c. e-po-ihyaɣot-a.
                     M-ABSC-run.away.2SG.UNDERGOER-EXTENDED
                     'You are running.' (male addressee)
               d. u-po-ihyaɣot-a.
                     F-ABSC-run.away.2SG.UNDERGOER-EXTENDED
                     'You are running'. (female addressee)
               e. e-p-ø-ihot-a.
                     M-ABSC-3SG.ACTOR-run.away.3SG.UNDERGOER-EXTENDED
                     'He is running (here where you are not looking).'
               f. u-p-ø-ihot-a.
                     F-ABSC-3SG.ACTOR-run.away.3SG.UNDERGOER-EXTENDED
                     'She is running (here where you are not looking).'

The inflected forms of Tayap in (11a–f) disclose not only a gender contrast in all three persons but also one that is identically coded across the three persons (i.e. /t/ for masculine and /k/ for feminine). It can also be seen that gender coding on the verb does not interact with gender marking on the subject pronoun.

Coastal Marind does not normally mark gender on the verb. However, it possesses a typologically uncommon grammatical category called the absconditive (ABSC), which assumes a mismatch between the speaker's and the listener's current focus of attention and encourages the listener to adopt the speaker's focus (Olsson 2019). As a matter of fact, the absconditive is accompanied by gender marking (/e/ for masculine and /u/ for feminine). As can be seen in (12), these gender markers occur in all persons (even though the second person is only rarely used).

It is remarkable that all the languages in Table 5 for which a full-person paradigm is available exhibit consistent gender marking. However, the lack of complete person paradigms in many grammars makes it impossible to determine how many inconsistent languages are included in Table 5. Inconsistent languages certainly exist. These are the languages lacking first-person gender marking. Two of them are Nepali (Indic; Acharya 1991) and Tunica (Tunica; Haas 1946), which distinguish

551

	Language	Macroarea	Family	Genus	Source
1	Ampeeli	Papunesia	Trans-New Guinea	Angan	Lloyd 1973
2	Barupu	Papunesia	Skou	Warupu	Corris 2005
3	Chaldean	Eurasia	Afro-Asiatic	Semitic	Sara 1974
4	Chipaya	South America	Uru-Chipaya	Uru-Chipaya	Cerrón-Palomino 2006
5	Coastal Marind	Papunesia	Anim	Marind	Olsson 2019/p.c.
6	Cubeo	South America	Tucanoan	Tucanoan	Chacon 2012
7	Culina	South America	Arauan	Arauan	Dienst 2014
8	Djingili	Australia	Mirndi	Djingili	Pensalfini 2004
9	Goajiro	South America	Arawakan	Carribean Arawakan	Ramirez Gonzales 1995
10	Hinuq	Eurasia	Nakh-Daghestanian	Avar-Andic-Tsezic	Forker 2013
11	Koasati	North America	Muskogean	Muskogean	Haas 1944
12	Korana	Africa	Khoe-Kwadi	Khoe-Kwadi	Meinhoff 1930
13	Kryz	Eurasia	Nakh-Daghestanian	Lezgic	Authier 2009
14	Kurukh	Eurasia	Dravidian	Northern Dravidian	Ekka 1972
15	Mehweb	Eurasia	Nakh-Daghestanian	Lak-Dargwa	Daniel 2019/p.c.
16	Molala	North America	Penutian	Molala	Pharris 2006
17	Mosetén	South America	Mosetenan	Mosetenan	Sakel 2002
18	Motuna	Papunesia	East Bougainville	East Bougainville	Onishi 2012
19	Piaroa	South America	Sáliban	Piaroa	Mosonyi 2002
20	Selknam	South America	Chon	Chon Proper	Rojas Berscia 2014
21	Sidaama	Africa	Afro-Asiatic	Highland East Cushitic	Kawachi 2007
22	Tayap	Papunesia	Tayap	Tayap	Kulick & Terrill 2019
23	Tsova Tush	Eurasia	Nakh-Daghestanian	Nakh	Helmbrecht 1996
24	Waunana	South America	Choko	Choko	Mejía Fonnegra 2000
25	Yana	North America	Hokan	Yana	Sapir 1949
26	Yucuna	South America	Arawakan	Inland Northern Arawakan	Schauer & Schauer 2000

Table 5
Non-Indo-European languages with first-person singular gender on verbs (N = 26).

between masculine and feminine forms in the third- and second-person present tense forms, though not in first-person forms. This is the same pattern that emerged in Semitic (see Section 3.4.1). It may be tentatively concluded that there is a tendency in verb paradigms to develop consistent gender marking strategies. Exactly how strong this tendency is remains to be worked out.

3.4.4 *Interim summary*

Participles have been identified as a preferred locus of gender marking on verbs. They code gender consistently across persons in all languages of the sample. Slavic languages are somewhat more likely to code gender on past tense than on present tense forms. When verb forms have first-person gender, they show consistent marking. When gender is inconsistently coded, Siewierska's person hierarchy is respected.

4. THEORETICAL DISCUSSION

Four domains of gender marking have been surveyed with an eye to assessing the likelihood of first-person singular gender as well as the likelihood of gender interacting with person. The general result is that while first-person gender marking may occur in all four domains, its frequency of occurrence varies from one domain to another. In line with the Redundancy Hypothesis, first-person gender is highly unlikely to occur on personal pronouns and possessive determiners.

At the same time, the Redundancy Hypothesis alone cannot account for the empirical data. It has to be supplemented with the Identity Hypothesis, which is required to handle not only the occurrence of first-person gender-specific pronouns but also the connotations that these pronouns may bear. Hence, both hypotheses are necessary for a comprehensive account of the data.

A comparison of the two hypotheses suggests that the Redundancy Hypothesis is considerably stronger than the Identity Hypothesis. Gender neutrality in first-person pronouns is much more prevalent than gender specificity among the languages of the world (30 as against 1720 languages). It seems that conciseness of form and simple paradigms are widely preferred to longer forms and more complex paradigms. This conclusion makes good sense in view of the high token frequency of pronouns and morphological gender markers in many gendered languages.

To assess paradigm consistency in personal pronouns generally, we return to Siewierska's (2013) analysis of the interaction of person and gender. Her sample includes 123 gendered languages of which not a single language possesses a consistent gender paradigm in the singular. Thus, consistency in personal-pronoun paradigms plays virtually no role in gendered languages.

Gender marking in possessive determiners has been less thoroughly investigated from a typological perspective. However, there is reason to believe that the percentages do not differ widely from those for the personal pronouns. Fewer

553

languages code gender in possessives than in personal pronouns (Berg 2020). Furthermore, as shown above, first-person gender is less common in possessives than in personal pronouns. The opportunity of consistent gender marking is therefore severely limited. It may accordingly be concluded that paradigm consistency in possessive determiners is a highly unlikely option.

The results for personal pronouns and possessive determiners contrast sharply with those for predicative adjectives and verbs. To the extent that the questionnaire sample and the set of additional languages provide a first approximation to the actual patterns, it may be argued that consistent gender marking is the rule in adjectives and possibly the preferred option in verbs. Whatever the true extent of consistent gender marking, it is certainly much higher on verbs and adjectives than in pronouns and determiners.

How do we account for the preference for gender marking consistency in verbs and adjectives and the preference for paradigm inconsistency in personal pronouns and possessive determiners? How can different components of the same language vary so massively in their gender marking strategies? Our point of departure is the distinction between two types of agreement variously termed grammatical vs. anaphoric (Bresnan & Mchombo 1987), grammatical vs. pragmatic (Wechsler & Zlatić 2000), lexical vs. semantic (Kraaikamp 2017), or lexical vs. referential (Dolberg 2019). I follow van Rijn's (2016a) lead in distinguishing between agreement and referential markers. Gender inflection on adjectival and verbal targets is a prototypical case of agreement marking. By contrast, personal pronouns and possessive determiners are referential markers because they may refer directly. It bears emphasizing that the distinction between referential and grammatical markers is not a categorical one. As van Rijn (2016b) points out, morphemes may vary along a scale from more referential to more grammatical. While personal pronouns and possessive determiners are certainly less referential than nouns, they are clearly more referential than agreement markers. Despite these gradient differences, the relevant markers will be named referential and grammatical units, for short.

It will be recalled that adjectives and verbs exhibit a good deal of paradigm consistency, whereas personal pronouns and possessive determiners are highly inconsistent in their gender marking across persons. In light of the above distinction, it may be argued that grammatical units show a high paradigm consistency, whereas referential units show a low paradigm consistency.

Why is it that the pressures towards consistency differ so vastly for the two types of units? Referential and grammatical units possess distinct properties. Grammatical units are integrated into a closed system that is largely immune to the contingencies of the outside world and therefore in a position to develop principles of its own. This 'remoteness' grants the same status to the members of a paradigm. Because these members are subject to the same constraints, gender marking may occur across the board.

Contrary to grammatical units, referential units form an open system. By their very nature, they are sensitive to the way the extralinguistic world is perceived and categorized by language users. This sensitivity opens the gates for a host of factors

554

that function to introduce variation. Two such factors are economy and redundancy. Since the sex of the discourse participants is obvious, there is no need to go to the trouble of creating gender-specific first- (and second-) person variants when a single form does the trick. The net effect is that referential units may differ in their status and enjoy a certain degree of autonomy even though they are members of the same paradigm. As a result, some persons may be gender-marked and others not.

It may be concluded that different person paradigms are subject to different organizational principles. Some paradigms have a tight fit, and others have a loose fit. Which type of fit is appropriate depends on the nature of the units to be organized. Grammatical units are tightly organized. They can develop consistent paradigms because their self-enclosed nature wards off influences from the extra-linguistic world. This self-enclosure is so strong that even intra-linguistic factors remain without effect. Gender agreement marking was found to be insensitive to whether the personal pronoun is gender-specific or gender-neutral. This insensitivity may lead to double marking, especially in the third person. This tolerance towards double marking may be argued to spring from the self-enclosed organization of grammatical units. In contrast, referential units are more loosely organized. They are geared to accommodate diverse, non-grammatical influences that may vary in strength for the different persons. As a consequence, referential units fail to develop consistent paradigms as far as the expression of gender is concerned.

As mentioned before, the distinction between referential and grammatical units is a matter of more-or-less rather than all-or-none. This gradient perspective permits us to accommodate gradient empirical effects. Gradience goes both ways: grammatical markers, which are usually consistent, may also show some inconsistency, and referential markers, which are usually inconsistent, may also show some consistency. For example, the Semitic languages were found to display some inconsistency in their verbal paradigms (see Section 3.4.1), whereas a low percentage of languages were found to display consistent gender marking in their personal-pronoun paradigms (see Section 3.1). These findings suggest that there is a weak tendency towards consistency in the paradigms of referential units as well as a weak tendency in the paradigms of grammatical units to be sensitive to such real-world effects as communicative redundancy. Notwithstanding major differences between referential and grammatical markers, some principles of gender marking are identical in both types of units. In particular, both (by and large) respect Siewierska's person hierarchy. For instance, inconsistent gender coding in the second person, but not in the first, is observed in both personal pronouns and verbs. The same two word classes hardly ever employ the reverse coding strategy. All this goes to show that referential and grammatical units are subject to similar constraints even though the strength of these constraints may differ immensely.

Both referential and grammatical units have hitherto been treated in an undifferentiated fashion. Clearly, there are additional effects impinging on paradigm consistency that are not covered by the distinction between referential and grammatical units. I have little to say about personal pronouns and possessive determiners. These two classes appear to be similarly reluctant to code gender

consistently. Whether there are (minor) differences between the two sets in their propensity for paradigm inconsistency is difficult to ascertain on the basis of a rather low number of relevant languages.

Among the grammatical markers, there are noticeable differences between adjectives and verbs. However, these differences do not result from their status as different word classes but rather from the fact that verbs tend to have person markers, whereas adjectives lack them. The simple observation is that full consistency emerges in the absence of person indices. While this result might seem unsurprising, it was not a foregone conclusion. The fact that there is gender agreement between the subject and the predicative adjective is unequivocal evidence of a grammatical relationship between the two. This relationship could theoretically form the basis for a person effect of the subject on the predicative adjective. However, there was no such effect. Its absence almost certainly results from the absence of person marking on the adjective. If there are no person markers, there is no sensitivity to person. And because there is no sensitivity to person, redundant coding sees the light of day. Ceteris paribus, the same explanation holds for participles. As non-finite forms, they are insensitive to person effects and hence generate fully consistent verb paradigms.

Finally, the claim that grammatical units bring about a higher degree of paradigm consistency than referential units embodies a notable prediction. Since paradigm consistency implies first-person gender, languages are predicted to be more likely to code first-person gender on grammatical than on referential units. The language sample on which the present study is based bears out this prediction: there are three times as many languages marking first-person gender on grammatical units as languages marking first-person gender on referential units. Specifically, when the four languages straddling the fence are set aside, 82 languages of the former type accompany 28 languages of the latter type. This is a highly conservative count because the 28-language sample cannot be substantially increased, while the 82-language sample presumably can. Recall that only those languages were taken into account for which explicit information on first-person gender was available. To conclude, first-person gender marking on grammatical units is 'easier' than first-person gender marking on referential units.

5. Conclusion

Contrary to received wisdom according to which first-person gender is a rarissimum among the languages of the world, this work has shown that, after all, first-person gender is not that infrequent. While it is true that only a limited number of languages make a gender distinction in first-person singular pronouns, a good number of languages introduce first-person gender 'through the back door'. What are the 'doors' through which first-person gender may enter grammatical systems? In view of the powerful Redundancy Hypothesis, it may be assumed that, metaphorically speaking, there are not many doors and the few existing ones are not wide open. It stands to reason that the initial entry point was referential. Nouns were used to distinguish

between female and male referents. In the grammaticalization process, this distinction may be preserved in personal pronouns – in particular, in the third person. For the first and second person, the Identity Hypothesis comes into play. The biological and social division of the sexes may create differential identities that are reflected in different ways in which male and female speakers refer to themselves or address others. However, the data presented in this article suggest that this is not the main door through which first-person gender makes its way into linguistic systems.

According to the standard grammaticalization cline, independent words develop into bound morphemes, of which the development from pronouns to inflections is one instantiation (e.g. Siewierska 1999; Haspelmath 2011). This development is accompanied by a gradual loss of referentiality and the emergence and gradual strengthening of paradigms. The more grammatical a paradigm, the higher its degree of self-enclosure. The tighter the organization of a paradigm, the more it strives for consistency. Following Carstairs-McCarthy (1998), paradigms show a natural tendency towards consistency.

This theoretical background enables us to identify one way in which first-person gender penetrates the system. Provided gender is preserved in the grammatical-ization process, it may enter the paradigm at any one point, with the third person being the most likely candidate. Owing to the closely-knit structure of the paradigm, gender generalizes across persons by means of analogy. While it cannot be ruled out that gender reaches all persons simultaneously, this is rather unlikely because personal pronouns code gender very unevenly across persons.

There is another way in which first-person gender may arise. This route hinges on the dissociation of gender and person and depends only indirectly on paradigms. As documented above, adjectives and participles are cross-linguistic-ally prone to gender marking. In their attributive use, adjectives modify nouns, which lack a person paradigm. This entails the absence of person indices on adjectives. When adjectives are used predicatively, they are controlled by pro-nominal subjects with a person paradigm. However, this constellation does not impose a paradigmatic structure on them. They remain person-neutral, so when they are sensitive to gender and the subject is a first-person pronoun, they cannot help but express first-person gender. The account of the behaviour of participles is essentially along the same lines. As one sign of deranking (Shagal 2019), participles do without person indices. When participles are sensitive to gender and the subject is a first-person pronoun, they must express first-person gender. This mechanism appears to be so automatic that it is completely oblivious to communicative redundancy.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://doi.org/10.1017/S0022226723000191.

REFERENCES

Acharya, Jayaraj. 1991. *A descriptive grammar of Nepali and an analyzed corpus*. Washington, D.C.: Georgetown University Press.

Aikhenvald, Alexandra Y. & R.M.W. Dixon. 1998. Dependencies between grammatical systems. *Language* 74, 56–80.

Audring, Jenny. 2014. Gender as a complex feature. *Language Sciences* 43, 5–17.

Authier, Gilles. 2009. *Grammaire kryz: Langue caucasique d'Azerbaïdjan, dialecte d'Alik*. Paris: Peeters.

Barikeri, V.H. 1982. A descriptive analysis of Lambadi language spoken in Bijapur district. Ph.D. thesis, Karnatak University, Dharwad.

Behnstedt, Peter. 1985. *Die nordjemenitischen Dialekte*. Wiesbaden: Reichert.

Berg, Thomas. 2020. Nominal and pronominal gender: Putting Greenberg's Universal 44 to the test. *Language Typology and Universals* 73, 525–574.

Beyer, Stephen V. 1993. *The Classical Tibetan language*. Delhi: Fine Art Press.

Bhat, D.N.S. 2004. *Pronouns*. Oxford: Oxford University Press.

Boelaars, Jan. 1950. *The linguistic position of South-Western New Guinea*. Leiden: Brill.

Bond, Oliver & Marina Chumakina. 2016. Agreement domains and targets. In Oliver Bond, Greville G. Corbett, Marina Chumakina & Dunstan Brown (eds.), *Archi. Complexities of agreement in cross-theoretical perspective*, 43–76. Oxford: Oxford University Press.

Boyeldieu, Pascal. 1982. *Deux études laal (Moyen-Chari, Tchad)*. Berlin: Reimer.

Bresnan, Joan & Sam A. Mchombo. 1987. Topic, pronoun and agreement in Chichewa. *Language* 63, 741–782.

Camp, Elizabeth & Millicent Liccardi. 1967. Itonama. In Esther Matteson (ed.), *Bolivian Indian grammars. Vol. 2*, 257–352. Norman: Summer Institute of Linguistics of the University of Oklahoma.

Carlson, Terry. 1991. *Tainae grammar essentials*. Ukarumpa.

Carstairs-McCarthy, Andrew. 1998. Paradigmatic structure: inflectional paradigms and morphological classes. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*, 322–334. Oxford: Blackwell.

Cerrón-Palomino, Rodolfo. 2006. *El Chipaya o la lengua de los hombres del agua*. Lima: Editorial de la Pontificia Universidad Católica del Perú.

Chacon, Thiago Costa. 2012. The phonology and morphology of Kubeo. Ph.D. thesis, University of Hawaii.

Chadwick, Neil. 1975. *A descriptive study of the Djingili language*. Canberra: Australian Institute of Aboriginal Studies.

Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.

Corris, Miriam. 2005. A grammar of Barupu: a language of Papua New Guinea. Ph.D. thesis, University of Sydney.

Creissels, Denis. 2008. Person variation in Akhvakh verb morphology: functional motivation and origin of an uncommon pattern. *Language Typology and Universals* 61, 309–325.

Criswell, Linda & Beverly Brandup. 2000. Un bosquejo fonológico y gramatical del Siriano. In María Stella Gonzáles de Pérez (ed.), *Lenguas indígenas de Colombia*, 395–417. Santafé de Bogota: Instituto Caro y Cuervo.

Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press.

da Costa, Januacele Francisca. 1999. *Ya:thê, a última língua nativa no nordeste do Brasil*. Ph.D. thesis, Universidade Federal de Pernambuco, Recife, Bresil.

Daniel, Michael. 2019. Mehweb verb morphology. In Michael Daniel, Nina Dobrushina & Dmitry Ganenkov (eds.), *The Mehweb language*, 73–115. Berlin: Language Science Press.

De Cássia Borella, Cristine. 2000. *Aspectos morfosyntáticos da língua Aweti (Tupi)*. Campinas: Unicamp.

de Oliveira, Maria das Dores. 2006. Ofayé, a língua do povo de mel: Fonologia e gramática. Ph.D. thesis, Universidade Federal de Alagoas.

de Pury, Sybille. 2003. Vice versa. Le genre en garifuna. *Faits de Langues* 21, 155–161.

de Vries, Lourens. 2004. *A short grammar of Inanwatan, an endangered language of the Bird's Head of Papua New Guinea, Indonesia*. Canberra: Australian National University.

Dienst, Stefan. 2014. *A grammar of Kulina*. Berlin: de Gruyter Mouton.

Dobson, Rose. 1997. *Gramática prática com exercícios da língua Kayabi*. Cuiabá: Sociedade Internacional de Lingüística.

Doctor, Raimond. 2004. *A grammar of Gujarati*. Munich: Lincom.

Dolberg, Florian. 2019. *Agreement in language contact*. Amsterdam: Benjamins.

Donohue, Mark & Lila San Roque. 2004. *I'saka. A sketch grammar of a language of north-central New Guinea*. Canberra: The Australian National University.

Dryer, Matthew S. & Martin Haspelmath. 2013. (eds.): *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Dunn, Michael. 2015. Gender determined dialect variation. In Greville G. Corbett (ed.), *The expression of gender*, 39–67. Berlin: de Gruyter Mouton.

Dwivedi, Amitabh Vikram. 2012. *A grammar of Hadoti*. Munich: Lincom.

Dwivedi, Amitabh Vikram. 2013. *A grammar of Bhadarwahi*. Munich: Lincom.

Dye, Melody, Petar Milin, Richard Futrell & Michael Ramscar. 2017. A functional theory of gender paradigms. In Ferenc Kiefer, James P. Blevins & Huba Bartos (eds.), *Perspectives on morphological organization*, 212–239. Leiden: Brill.

Ekka, Francis. 1972. Men's and women's speech in Kurux. *Linguistics* 81, 25–31.

Elfenbein, Josef. 1972. Waneci. *Afghanistan: Historical and Cultural Quarterly* 25, 29–72.

Forchheimer, Paul. 1953. *The category of person in language*. Berlin: Walter de Gruyter.

Forker, Diana. 2013. *A grammar of Hinuq*. Berlin: de Gruyter Mouton.

Forker, Diana. 2020. *A grammar of Sanzhi Dargwa*. Berlin: Language Science Press.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–113. Cambridge, MA: MIT Press.

Grice, Herbert P. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics. Vol. 3: Speech acts*, 41–58. New York: Academic Press.

Güldemann, Tom. 2004. Reconstruction through 'de-construction'. The marking of person, gender, and number in the Khoe family and Kwadi. *Diachronica* 21, 251–306.

Gusain, Lakhan. 2001. *Shekhawati*. Munich: Lincom.

Haas, Mary R. 1944. Men's and women's speech in Koasati. *Language* 20, 142–149.

Haas, Mary R. 1946. A grammatical sketch of Tunica. In Harry Hoijer & Cornelius Osgood (eds.), *Linguistic structures of native America*, 337–366. New York: The Viking Fund.

Haig, Geoffrey. 2019. Northern Kurdish. Kurmanjî. In Geoffrey Haig & Geoffrey Khan (eds.), *The languages and linguistics of Western Asia*, 106–158. Berlin: de Gruyter Mouton.

Haiman, John. 1983. Iconic and economic motivation. *Language* 59, 781–819.

Harder, Andreas. 1988. Laut- und Formenlehre der Mundart von Ripatransone. Ph.D. thesis, University of Kiel.

Haspelmath, Martin. 2011. The gradual coalescence of 'words' in grammaticalization. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 342–355. Oxford: Oxford University Press.

Haspelmath, Martin. 2013. Argument indexing: a conceptual framework for the syntactic status of bound person forms. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries*, 197–226. Berlin: de Gruyter Mouton.

Haurholm-Larsen, Steffen. 2016. *A grammar of Garifuna*. Ph.D. thesis, University of Bern, Switzerland.

Hellenthal, Anne-Christie. 2010. *A grammar of Sheko*. Utrecht: LOT.

Helmbrecht, Johannes. 1996. The syntax of personal agreement in East Caucasian languages. *Language Typology and Universals* 49, 127–148.

Hinds, John. 1986. *Japanese*. London: Routledge.

Ibriszimov, Dimitr & Guillaume Segerer. 2004. (eds.): *Systèmes de marques personelles en Afrique*. Paris: Peeters.

Jenny, Mathias & San San Hnin Tun. 2016. *Burmese. A comprehensive grammar*. London: Routledge.

Jung, Ingrid. 1989. *Grammatik des Paez: Ein Abriss*. Ph.D. thesis, Universität Osnabrück, Germany.

Kawachi, Kazuhiro. 2007. A grammar of Sidaama (Sidamo), a Cushitic language of Ethiopia. Ph.D. thesis, State University of New York at Buffalo.

Khalilova, Zaira. 2009. *A grammar of Khwarshi*. Utrecht: LOT.

Khan, Geoffrey. 1999. *A grammar of Neo-Aramaic. The dialect of the Jews of Arbel*. Leiden: Brill.

Khan, Geoffrey. 2004. *The Jewish Neo-Aramaic dialect of Sulemaniyya and Halabja*. Leiden: Brill.

Khan, Geoffrey. 2008a. *The Neo-Aramaic dialect of Barwar. Vol. 1*: *Grammar*. Leiden: Brill.

Khan, Geoffrey. 2008b. *The Jewish Neo-Aramaic dialect of Urmi*. Piscataway: Gorgias.

Kibrik, Alexandr. 1996. *Godoberi*. Munich: Lincom.

559

Kinch, Rodney & Pamela Kinch. 2000. El Juruti. In María Stella Gonzáles de Pérez (ed.), *Lenguas indígenas de Colombia*, 469–487. Santafé de Bogota: Instituto Caro y Cuervo.

Kraaikamp, Margot. 2017. The diachrony of semantic gender agreement: Findings from Middle Dutch. *Journal of Germanic Linguistics* 29, 259–297.

Kulick, Don & Angela Terrill. 2019. *A grammar and dictionary of Tayap*. Berlin: de Gruyter Mouton.

Laycock, D.C. 1965. *The Ndu language family (Sepik District, New Guinea)*. Canberra: Australian National University.

Lehr, Rachel. 2014. *A descriptive grammar of Paskai: The language and speech community of Darrai Nur*. Ph.D. thesis, University of Chicago, IL.

Liljegren, Henrik. 2016. *A grammar of Palula*. Berlin: Language Science Press.

Lloyd, Richard G. 1973. The Angan language family. In Karl J. Franklin (ed.), *The linguistic situation in the Gulf District and adjacent areas, Papua New Guinea*, 33–110. Canberra: Australian National University.

Mansfield, John. 2019. *Murrinhpatha morphology and phonology*. Berlin: de Gruyter Mouton.

Mayerthaler, Willi. 1988. *Morphological naturalness*. Ann Arbor, MI: Karoma.

Meinhof, Carl. 1930. *Der Koranadialekt des Hottentottischen*. Berlin: Reimer.

Mejía Fonnega, Gustavo. 2000. Presentación y descripción fonológica y morfosyntáctica del Waunana. In María Stella Gonzáles de Pérez (ed.), *Lenguas indígenas de Colombia*, 85–96. Santafé de Bogota: Instituto Caro y Cuervo.

Miranda, Rocky V. 2007. Konkani. In George Cardona & Dhanesh Jain (eds.), *The Indo-Aryan languages*, 803–846. London: Routledge.

Moore, Bruce R. 1979. *Metodo para aprender el idioma colorado (gramatica pedagogica). Vol. 1*. Quito: Instituto Linguistico de Verano.

Morgenstierne, Georg. 1945. Notes on Shumashti. *Norsk Tidsskrift for Sprogvidenskap* 132, 239–273

Mosonyi, Esteban Emilio. 2002. Elementos de gramática piaroa: algunas consideraciones sobre sus clases nominales. *Boletín de Lingüística* 18, 71–103.

Mutzafi, Hezy. 2004. *The Jewish Neo-Aramaic dialect of Koy Sanjaq*. Wiesbaden: Harrassowitz.

O'Hagan, Zachary. 2011. Proto-Omagua-Kokama: Grammatical sketch and prehistory. M.A. thesis, University of California at Berkeley.

Oates, W. & L. Oates. 1968. *Kapau pedagogical grammar*. Canberra: Australian National University.

Olsson, Bruno. 2019. The absconditive revealed: attention alignment in the grammar of Coastal Marind. *Open Linguistics* 5, 136–155.

Onishi, Masayuki. 2012. *A grammar of Motuna*. Munich: Lincom.

Pandharipande, Rajeshwari V. 1997. *Marathi*. London: Routledge.

Parihar, Ravi & Amitabh Vikram Dwivedi. 2019. *A grammar of Sarazi*. Munich: Lincom.

Pensalfini, Rob. 2004. Towards a typology of configurationality. *Natural Language and Linguistic Theory* 22, 359–408.

Pharris, Nicholas J. 2006. Winuunsi Tm Talapaas: A grammar of the Molala language. Ph.D. thesis, University of Michigan.

Plank, Frans & Wolfgang Schellinger. 1997. The uneven distribution of genders over numbers: Greenberg Nos. 37 and 45. *Linguistic Typology* 1, 53–101.

Ramirez Gonzales, Rudecindo. 1995. *Morfologia del verbo en Wayunaiki*. M.A. thesis, Universidad de los Andes, Santafe de Bogota.

Rebeiro, Eduardo Rivail. 2012. *A grammar of Karajá*. Ph.D. thesis, University of Chicago, IL.

Rivet, P., P. Kok & C. Tastevin. 1924-25. Nouvelle contribution à l'étude de la langue Maku. *International Journal of American Linguistics* 3, 133–192.

Rogers, Phillip G. & Stefan Th. Gries. 2022. Grammatical gender disambiguates syntactically similar nouns. *Entropy* 24, e24040520.

Rojas Berscia, Luis Miguel. 2014. A heritage reference grammar of Selk'nam. M.A. thesis, Radboud University Nijmegen.

Rose, Françoise. 2013. Le genre du locuteur et de l'allocutaire dans les systèmes pronominaux: Genre grammatical et indexicalité du genre. *Bulletin de la Société Linguistique de Paris* 58, 381–417.

Rose, Françoise. 2015. On male and female speech and more: categorical gender indexicality in indigenous South American languages. *International Journal of American Linguistics* 81, 495–537.

Rose, Françoise. 2018. A typology of languages with genderlects and grammatical gender. In Sebastian Fedden, Jenny Audring & Greville G. Corbett (eds.), *Non-canoninal gender systems*, 211–246. Oxford: Oxford University Press.

Sakel, Jeanette. 2002. Gender agreement in Mosetén. In Mily Crevels, Simon van de Kerke, Sérgio Meira & Hein van der Voort (eds.), *Current studies on South American languages*, 287–305. Leiden: Research School of Asian, African and Amerindian Studies.

Sands, Bonnie. 2013. Hadza. In Rainer Vossen (ed.), *The Khoesan languages*, 107–123. New York: Routledge.

Sapir, Edward. 1949. Male and female forms of speech in Yana. In David G. Mandelbaum (ed.), *Selected writings of Edward Sapir*, 206–212. London: Cambridge University Press.

Sara, Salomon I. 1974. *A description of Modern Chaldean*. The Hague: Mouton.

Schauer, Stanley & Junia Schauer. 2000. El Yucuna. In María Stella Gonzáles de Pérez (ed.), *Lenguas indígenas de Colombia*, 515–532. Santafé de Bogota: Instituto Caro y Cuervo.

Schmidt, P.W. 1919a. *Die Personalpronomina in den australischen Sprachen*. Wien: Alfred Hölder.

Schmidt, P.W. 1919b. *Die Gliederung der australischen Sprachen*. Wien: Mechitharisten.

Shagal, Ksenia. 2019. *Participles. A typological study*. Berlin: de Gruyter Mouton.

Sieg, Emil & Wilhelm Siegling. 1931. *Tocharische Grammatik*. Göttingen: Vandenhoek & Ruprecht.

Siewierska, Anna. 1999. From anaphoric pronoun to grammatical agreement marker: Why objects don't make it. *Folia Linguistica* 33, 225–251.

Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.

Siewierska, Anna. 2013. Gender distinctions in independent personal pronouns. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Silva, Léia de Jesus. 2005. Aspectos da fonologia e da morfologia da língua rikbáktsa. M.A. thesis, Universidade de Brasília.

Smyth, David. 2002. *Thai. An essential grammar*. New York: Routledge.

Stenzel, Kristine S. 2013. *A reference grammar of Kotiria (Wanano)*. Lincoln: University of Nebrasca Press.

Tal, Shira & Inbal Arnon. 2022. Redundancy can benefit learning: evidence from word order and case marking. *Cognition* 224, 105055.

Tsehay Taye, Kibebe. 2015. Documentation and grammatical description of Chabu. Ph.D. thesis, Addis Abeba University.

Vallejos, Rosa. 2016. *A grammar of Kukama-Kukamiria: A language from the Amazon*. Leiden: Brill.

Vamling, Karina & Revaz Tchantouria. 1991. Sketch of the grammar of Kubachi. *Lund University Working Papers in Linguistics* 38, 221–236.

Van Rijn, Marlou. 2016a. Locus of marking typology in the possessive NP: A new approach. *Folia Linguistica* 50, 269–327.

Van Rijn, Marlou. 2016b. The grammaticalization of possessive person marking: A typological approach. *Transactions of the Philological Society* 114, 233–276.

van Staden, Miriam. 2000. Tidore: A linguistic description of a language of the North Moluccas. Ph.D. thesis, Universiteit Leiden.

Vanhove, Martine. 2017. *Le bedja*. Leuven: Peeters.

Varma, Siddheshwar. 1948. *The Bhalesi dialect*. Calcutta: Royal Asiatic Society of Bengal.

Vlcek, Nathalie Pires. 2016. Documentação linguistica Utapinopona-Tuyuca: aspectos fonológicos e morfológicos. Ph.D. thesis, Universidade Federal do Rio de Janeiro.

Voorhoeve, C.L. 1975. The central and western areas of the Trans-New Guinea phylum. In Stephen A. Wurm (ed.), *New Guinea area languages and language study. Vol. 1*, 345–459. Canberra: Australian National University.

Wali, Kashi & Omkar N. Koul. 1997. *Kashmiri. A cognitive-descriptive grammar*. London: Routledge.

Waltz, Carolyn & Nathan Waltz. 2015. *Notas sobre el piratapuyo*. SIL International.

Watuseke, F.S. 1991. The Ternate language. In Tom Dutton (ed.), *Papers in Papuan linguistics. Vol. 1*, 223–244. Canberra: Australian National University.

Wechsler, Stephen & Larisa Zlatić. 2000. A theory of agreement and its application to Serbo-Croatian. *Language* 76, 799–832.

Wedekind, Klaus, Charlotte Wedekind & Abuzeinab Musa. 2005. *Beja pedagogical grammar*. Asmara.

West, Birdie. 1980. *Gramatica popular del Tucano*. Bogotá: Instituto Lingüístico de Verano.

Author's address: *Department of English, University of Hamburg,*
                       Von-Melle-Park 6, 20146 Hamburg, Germany
                       thomas.berg@uni-hamburg.de