

Kierán Meinhardt 

Freie Universität Berlin Germany | kieran.meinhardt@fu-berlin.de

## ***‘Je’ est une autre: Gendered First-Person Singular Pronouns in Tocharian A and Yemeni Arabic Dialects***

**Abstract** This paper examines the typologically rare phenomenon of grammatical gender distinction in first-person singular pronouns, focusing on Tocharian A and Yemeni Arabic dialects. Neither language inherited this feature—it is absent from common Indo-European and common Semitic. Instead, sociolinguistic factors likely drove its emergence. In Tocharian A, it may stem from hypercorrection, while in Yemeni Arabic dialects, it reflects competition between local and supralocal forms. Some evidence suggests a contact influence, possibly from a Himyaritic substrate. Ultimately, both cases highlight how common sociolinguistic tendencies can lead to the emergence of unusual grammatical structures.

**Keywords** typology of pronouns, historical Arabic dialectology, Yemen, Indo-European, Tocharian A, sociolinguistics, grammatical gender, substrate, Himyaritic

*Je est un autre.*

Arthur Rimbaud, French poet,  
spent part of 1880 in Aden

### **1 Typological prelude**

According to data from the typology database *Grambank* (Skirgård et al. 2023, Feature GB197), 30 out of 2186 languages studied—just about 1.37%—express a gender distinction in the first-person pronoun. Even this already small number overstates the phenomenon as studied here, since it includes languages in which gender is marked only in non-singular forms. Slovenian, for instance, distinguishes between plural *mî* vs. *mê* for mixed-gender or all-male groups versus all-female groups (Corbett 1991: 129; Cysouw 2003: 319; Plank and Schellinger 1997: 72).



Functionally, this asymmetry makes perfect sense:

Assuming that the speaker is visible to his or her addressees, why should he or she bother to tell them his or her own gender? In the 1st person non-singular, on the other hand, reference is being made by the speaker to others not necessarily visible rather than to only him- or herself, and maybe someone might be obliged to know of which gender these associates of his or hers are. (Plank and Schellinger 1997: 65)<sup>1</sup>

In fact, I was able to verify a gender distinction in the singular pronoun for only 12 of the languages surveyed in *Grambank*—just 0.55%.<sup>2</sup> Curiously, neither of the two languages at the focus of this study appears on that list: Tocharian A (section 2) and the cluster of Arabic dialects spoken in Yemen’s western coastal regions (section 3). These two languages were selected because, unlike most others, they come from well-researched language families with a great deal of historical depth. The following sections compare how this rare phenomenon emerges in each case from three perspectives:

1. From which ancestral forms were the different pronoun forms inherited or newly derived? (2.1 / 3.1)
2. What social conditions or general sociolinguistic trends may have encouraged the emergence of this distinction? (2.2 / 3.2)
3. Are there indications that contact with adstrate or substrate languages contributed to the development of gender distinction? (2.3 / 3.3)

	Tocharian A	Arabic dialects of the Tihāmeḥ
1s	<i>näṣ</i> m. / <i>ñuk</i> f.	<i>ʾana</i> m. / <i>ʾani</i> f.
2s	<i>tu</i>	<i>ʾanta</i> m. / <i>ʾanti</i> f.
3s	<i>sām</i> m. / <i>sām</i> f. / <i>tām</i> n.	<i>hū</i> m. / <i>hī</i> f.

<sup>1</sup> See also Cysouw (2003: 319): ‘It is fairly uncommon for reference to the speaker to distinguish gender’.

<sup>2</sup> These are Shabo (m. *tijj*, f. *taja*) (Kibebe Tsehay Taye 2015: 162), Hadza (m. *ʾono*, f. *ʾo-noko*) (Sands 2013: 112) and Laal (m. *já*, f. *ji*) (Lionnet 2010: 9) in Africa, as well as Thai (m. *phǒm*, f. *(di)chán*) (Smyth 2002: 39), Sô (m. *ŋkua?*, f. *ŋkɔ?*) (Lerthirunwong 1980: 139), Tengger Javanese (m. *(r)eyang*, f. *isun*) (Connors 2008: 43), and Japanese (m. *boku*, f. *atashi*, common *wata(ku)shi*) (Hinds 1986: 239) in East Asia, and Rikbaktsa (m. *uta*, f. *ikra*) (de Jesus Silva 2011: 52), Ofayé (m. *a*, f. *agʷ*) (das Dores de Oliveira 2006: 197), Fulniô (m. *owe*, f. *oso*) (da Costa 1999: [161]), Tsafiki (m. *la*, f. *chike/tse*) (Dickinson 2002: 65), and Cocama-Cocamilla (m. *ta*, f. *etse/tsa*) (Vallejos Yopán 2010: 201) in South America.

	Tocharian A	Arabic dialects of the Tihāmeḥ
1p	<i>was</i>	<i>ʾihna</i>
2p	<i>yas</i>	<i>ʾantun m. / ʾantin f.</i>
3p	<i>cem m. / tom f.</i>	<i>hun m. / hin f.</i>

Table 1. The independent personal pronouns in Tocharian A (Krause and Thomas 1960: 162) and the Arabic dialects of the Yemeni coastal plain (Tihāmeḥ) based on informants from around Bāḡil (Diem 1973: 65, 68)

## 2 Tocharian A

The two Tocharian languages (A and B) form an independent branch of the Indo-European language family. They are attested on written records from the 6th–8th centuries in the northwest of present-day China (Xinjiāng). Tocharian A (also called East Tocharian, henceforth TA) originates from the area around the Turfan and Qarašahr oases on the edge of the Taklamakan desert (Krause and Thomas 1960: 37).<sup>3</sup>

The majority of texts in Tocharian are of Buddhist content, often translated from Sanskrit or Middle Indic originals. The following excerpt, taken from a fragment of the *Ṣaḍdanta-Jātaka* (Carling, Pinault, and Malzahn 2014), offers a compelling illustration of grammatical gender distinctions in the first-person singular pronoun. In this passage, King Mahendrasena addresses his daughter Bhadrā, whom he intends to marry off—unsure whom to choose as a suitable husband for her.

(1) *caṣ nāṣ [...] tṗār ṣtānikaṃ lymā āleyaṃ ṣanweṃyo* (A 66 a1)

‘[...] I was sitting in the palace above with [my] cheeks in the palm of [my] hand.’

Ultimately, Bhadrā resolves the matter herself, choosing a husband according to her own will:

(2) *(tä)my(o) ṅuk p(e)nu [...] ṣñi mänwā pats yāmmār* (A 66 b1)

‘(There)fore I, too, will choose a husband for myself [...], according to [my] own will’

The contrast between the forms *nāṣ* and *ṅuk*—the former masculine, the latter feminine—reflects a switch in grammatical gender aligned with the speaker’s identity.

<sup>3</sup> See Fortson (2010) for a short introduction to the Tocharian languages.

## 2.1 Historical phonology

The Tocharian system presented in Table 1 distinguishes gender in the first-person singular and third-person pronouns. The language most closely related to TA, Tocharian B, as well as the other ancient Indo-European languages, make this distinction in the third-person pronouns only. A system like that of TA is typologically unique.<sup>4</sup>

It is highly unlikely that such a system was inherited, meaning it existed in the proto-stages—Proto-Tocharian and even Proto-Indo-European—vanished in all other Indo-European languages, and was only preserved in TA (Bross 2022: 86). The gender distinction in TA’s first-person singular is therefore an innovation. But what are the Proto-Indo-European (PIE) antecedents of these two-word forms?

The reconstruction of these forms remains, as summarized by Bross (2022: 80) in reviewing Adams, Winter, and Kim, ‘as thorny a thicket of morphology and phonology as one can find’ and results in ‘no satisfactory analysis’ of the forms, which are ‘scarcely less baffling today than they were a century ago’.

Nonetheless, the most widely accepted view should be presented here.<sup>5</sup> It suggests that the TA masculine form *näš* (which corresponds to the Tocharian B form for both genders) derives from a combination of the old accusative of the first person PIE *\*me* (cf. Latin *mē*, Greek *me*) and the particle PIE *\*ge* (cf. Greek *ge*, Slavic *že*),<sup>6</sup> while the TA feminine form *ñuk* results from a contamination of the old nominative of the first person PIE *\*eǵō* with the oblique (i.e., non-nominative) case forms, which in Tocharian begin with *ñ-*.

Schematically (taken from Bross 2022: 83):

PIE Acc. <i>*me-ge</i>	→	PToch. <i>*ñāśā</i>	→	TB <i>ñās</i> (m./f.), TA <i>näš</i> (m.)
PIE Nom. <i>*eǵō</i>	→	PToch. <i>*ñāku</i>	→	TA <i>ñuk</i> (f.)

Here, *ñ-* arises from the palatalization of *\*m* before the front vowel *\*e*, while *ś* reflects a palatalized *\*g* in the same environment. By contrast, the *k* in *ñuk* continues *\*g* in a non-palatalizing context.

<sup>4</sup> See Bross (2022: 88): ‘Genus in der 1. Person Singular des Personalpronomens ist an sich schon sehr außergewöhnlich, aber für die paradigmatische Struktur des Osttocharischen sind keine typologischen Parallelen bekannt’, because ‘if [gendered pronouns] are found in 1st, which is rare, they will normally also be there in 2nd and 3rd’ (Aikhenvald 2000: 252).

<sup>5</sup> See Bross (2022: 79–85) for an overview of all reconstructions.

<sup>6</sup> Cf. also German *mich*, which is derived from the same formation.

## 2.2 Sociolinguistics

In this reconstruction, the key question remains: how did the case distinction between nominative and accusative develop into a gender distinction between feminine and masculine? Given that TA has been extinct for over a thousand years and the surviving texts provide no direct evidence on the genesis of this gender difference, any explanation must remain speculative.

One possible explanation is that these forms reflect a generalization of tendencies towards different case usage in men's and women's speech. A parallel example is found in English, where speakers differ in their use of nominative and accusative pronouns after the copula, such as *It is I* vs. *It is me*. The former—more traditional—nominative form is considered a 'high-register variant', often characterized as an 'overly careful, priggish [...] way of talking—a style that tends to be stereotyped in the popular mind as a female characteristic' (Jasanoff 1989: 136). This aligns with sociolinguistic findings that women 'allowing for other factors such as social class, ethnic group and age, [...] on average use forms which more closely approach those of the standard variety [...] than those used by men' (Trudgill 2000: 70).

Thus, after the case distinction between nominative and accusative collapsed in TA pronouns, both forms may have coexisted in free variation. In men's speech, the innovative form prevailed—eventually becoming the masculine pronoun—while in women's speech, the conservative form was retained (Trudgill 2000: 77–78) and hypercorrectively extended even to accusative contexts.<sup>7</sup>

Additionally, Jasanoff (1989: 139) suggests that this grammaticalization process may have been influenced by contact with neighboring Tibeto-Burman languages, known for their extensive pronominal differentiation by gender and social register.

## 2.3 Language contact

According to Jasanoff (1989: 139), Tibetan distinguishes first-person singular pronouns by gender, using *k'o-wo* (m.) and *k'o-mo* (f.), while Burmese has politeness-based distinctions such as *cuñto* (m., lit. 'royal slave') and *cuñmá* (f., lit. 'female slave'). These connections with Tibeto-Burman have first been noticed by Hermann (1922: 309).

However, more recent studies indicate that the Tibetan distinction is based on social deixis rather than gender, and the Burmese forms function more as honorifics rather than true personal pronouns. Furthermore, no plausible scenario for sustained contact between Tocharian and Tibeto-Burman languages has been

---

<sup>7</sup> The tendency toward hypercorrection aligns with the observation that 'many societies seem to expect a higher level of adherence to social norms—better behaviour—from women than they do from men' (Trudgill 2000: 73).

demonstrated (Bross 2022: 86–87), but see Ivanov (2008: 588–89) for further linguistic parallels between Tocharian and Tibetan.

While it remains possible that areal factors or influence from an unknown language contributed to the emergence of gender distinction in TA, no concrete evidence supports this hypothesis (Bross 2022: 87).

### 3 Yemeni Arabic dialects

As early as the 19th century, gender distinction in the first-person singular pronoun was documented for Ḥaḍramaut Arabic (van den Berg 1886: 248–249).<sup>8</sup> However, more recent research has shown that this phenomenon is not limited to Ḥaḍramaut. Behnstedt (1985; 2016) found that the distinction extends across most of Yemen’s western coastal region of Yemen (regions B and C in figure 1), including major cities such as Ibb and Aden.<sup>9</sup>

One illustrative example comes from the Jewish dialect of Giblih, spoken south of Ibb (Shachmon 2022: 184). In this exchange, a girl is speaking to her father. Her use of the feminine-marked form of the first-person singular pronoun (*anī*) stands in contrast to her father’s masculine-marked form (*anā*):

(3) *anī ša-rwī-be yā-bah, mā ša-tiktaser*  
‘I will fill it [the pot] with water, dad, it will not break’

However, the pot breaks, and her father responds:

(4) *anā qulk<sup>q</sup>lik yā bintī, qulk<sup>q</sup>lik ša-tiksirī, mā samma‘tīnīš*  
‘I had told you, my daughter, I had told you you might break it, but you did not listen to me’

#### 3.1 Historical phonology

When Arabic dialects show deviations from the standard language in the pronominal system, it is usually the elimination of gender distinctions in the 2nd and 3rd person plural (Hachimi 2006: 161). The presence of an additional gender

<sup>8</sup> Brockelmann (1908 1: 297) attributed this distinction uniquely to Ḥaḍramaut Arabic among all Semitic languages.

<sup>9</sup> The same gender distinction is said to occur—spatially separated—in individual Shiite dialects in Bahrain (Isaksson 1999: 60). However, further information, evidence, and sources for this are lacking.

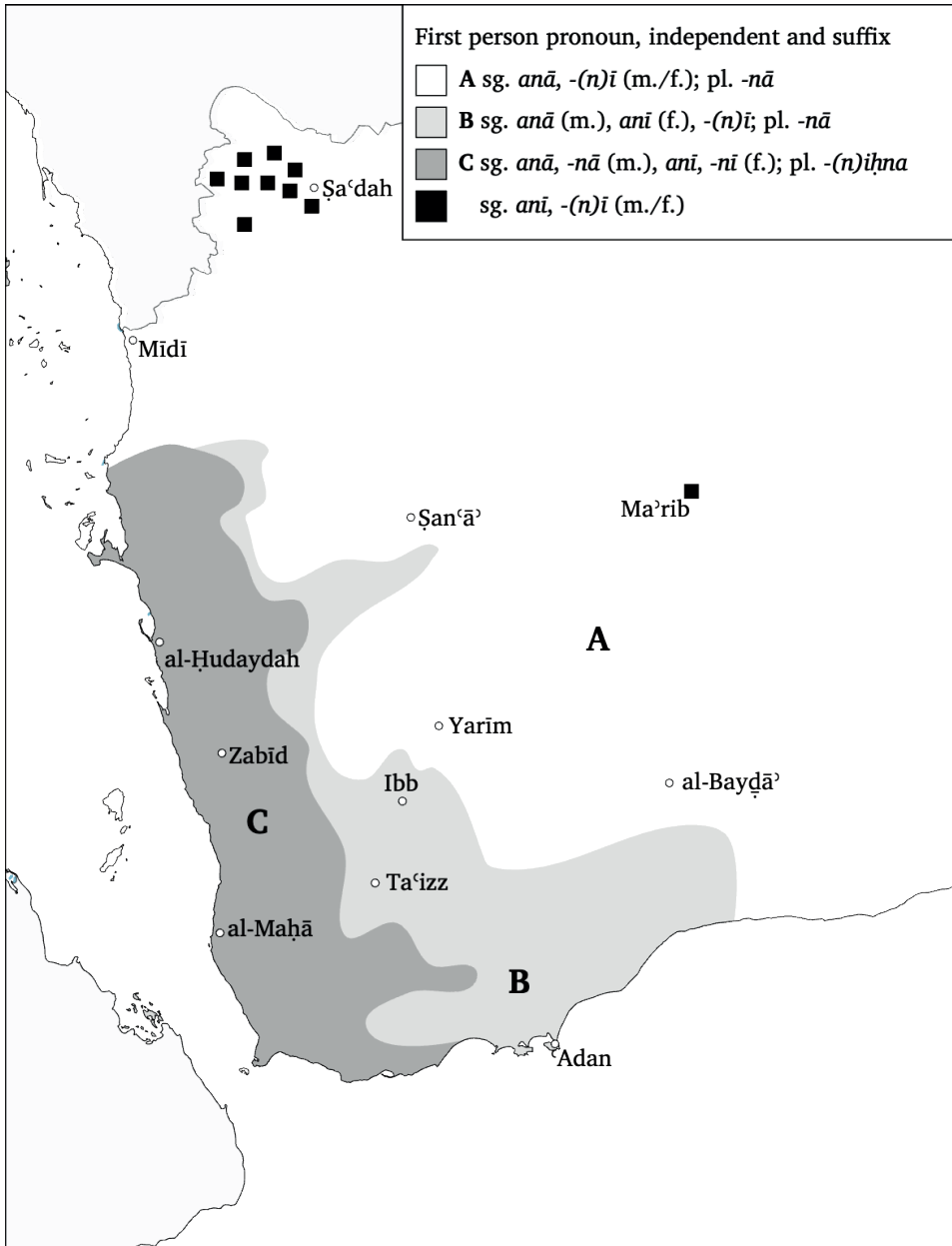


Figure 1. Distribution of gender distinction in the first-person singular pronoun in western Yemen (map based on Behnstedt and Woidich 2010: 151). Region A (white) has no gender distinction in the 1st person, region B (light grey) has a gender distinction in the independent pronoun but not in the suffix pronoun, region C (darker) has a gender distinction in both the independent pronoun and the suffix pronoun

distinction in the first-person singular pronoun in Yemeni dialects—considered among the most conservative varieties of Arabic (Behnstedt and Woidich 2010: 15, 24)—raises the question of whether this feature is inherited from Proto-Semitic or Proto-Afroasiatic, or whether it is a more recent innovation.

As already discussed in relation to Tocharian A (2.1), for such a typologically rare feature, it is more likely to be an innovation within the current area of occurrence. If the formal difference had been inherited from older Arabic, it would have had to have been eliminated in all dialects from Mauritania to Afghanistan before the beginning of written records, without leaving any traces. If it were inherited from Proto-Semitic or Proto-Afroasiatic, the question arises why no other Semitic or Afroasiatic language distinguishes between a masculine and feminine pronoun of the 1st person singular.<sup>10</sup>

Brockelmann (1908: 1:297)—followed by Diem (1973: 68) and Werbeck (2001: 64)—therefore proposed that this distinction arose through paradigmatic analogy:

$$ʔana:ʔanta = x:ʔanti$$

The idea is that speakers extended the pattern seen in the second-person singular—*ʔanta* (m.) vs. *ʔanti* (f.)—to the first person. Since *ʔana* resembles *ʔanta*, a new feminine form (*ʔani*) could have arisen by analogy to maintain gender symmetry across the paradigm.

### 3.2 Sociolinguistics

The ‘Arab-Islamic [...] notion of space dichotomy’, which assigns men to the public sphere and women to the private sphere (Sadiqi 2011), is reflected in differences between men’s and women’s speech (Behnstedt and Woidich 2010: 60).<sup>11</sup> However, these differences run counter to the tendency observed in TA (2.2), where women’s speech generally aligns more closely with the prestige variety. In the Arab world, ‘women who are denied education and/or travel will obviously not use more standard forms than men’ (Trudgill 2000: 73).<sup>12</sup> As a result, women often retain nonstandard forms, making female speech more conservative than

<sup>10</sup> See Takács (2011: 10) and Appleyard (2011: 44) for an overview of the personal pronouns of Afroasiatic languages.

<sup>11</sup> See Al-Wer (2014), Bassiouney (2020), and Vicente (2021) for examples.

<sup>12</sup> See also Isaksson (1999: 60): ‘the habit of the women referring to themselves with *ʔani* is regarded as an uneducated behaviour’.

male speech (Behnstedt and Woidich 2010: 60),<sup>13</sup> while men tend to adopt more standard-like forms.<sup>14</sup>

However, Modern Standard Arabic does not function as the prestige variety in this case, 'nor does it have a normative effect on the structure of variation in spoken Arabic in the core domains of phonology, morphology, and syntax' (Al-Wer 2014: 402). Instead, this role is primarily taken by supralocal forms, such as the dialect spoken in the capital city.

Although no dedicated sociolinguistic studies have been conducted on the *'ana* vs. *'ani* distinction, the above-mentioned trends allow for certain predictions: Since the first-person singular pronoun is *'ana* in both Standard Arabic and in the dialect of the capital, *Ṣan'ā'* (cf. Figure 1), the form *'ana* would be expected to become the gendered form used by men. Meanwhile, an older, local form *'ani*, which deviates from the standard, would be expected to stabilize in women's speech.

### 3.3 Language contact

For a distinction like *'ana* vs. *'ani* to become established through the sociolinguistic processes described above, both forms must have previously existed in free variation. This raises the question of chronology and the origin of the form *'ani*: where was it first present, and where did it come from?

Insights into the historical distribution of a linguistic form can often be derived from the shape of its isogloss (cf. Figure 1). In the case of the first-person singular pronoun, a funnel shape can be observed pointing towards *Ṣan'ā'*, extending from the flat coastal plain (Tihāmah) into the highlands. The extent appears to be primarily determined by valleys, meaning that Tihāmah residents originally settled in these easily accessible areas or migrated there. The key question is whether we are looking at a 'funnel'—that is, *'ana* (m.) and *'ani* (f.) advancing into the *'ana* region—or, conversely, whether two 'wedges' of *'ana* are penetrating into an area that originally featured both *'ana* and *'ani* (Behnstedt and Woidich 2010: 150). Below, I present three possible explanations.

---

<sup>13</sup> See also Isaksson (1999: 60). According to the 'Gender Paradox' noted by Labov (2006 2: 280), innovative forms—when there isn't a top-down prescribed standard—are actually more common among women than men. However, it seems that the form *'ani* aligns well with conservative Bedouin and pre-Arabic forms (see below), which may suggest that this 'Gender Paradox' doesn't fully apply in this particular context.

<sup>14</sup> Cf. Sadiqi (2011): 'one could probably say that the standard form of the language is usually felt to be more associated with men than with women'.

### Scenario 1: Bedouin Arabic

The origin of the feminine form *ʔani* could lie in the Bedouin dialects where the first-person singular pronoun appears as *ʔani/ʔane* with final *ʔimāla* (fronting) (Rosenhouse 2006: 263). This form is also found in Yemen in the dialects around *Ṣaḍah* and *Mārib*, where it is used for both genders (cf. Figure 1). The current spread could be explained by the hypothesis that the Bedouin form *ʔani* was once used throughout western Yemen but was gradually displaced by *ʔana*, the form used in dialects of northern Arabian origin.<sup>15</sup> As the form *ʔana* spread and gained prestige, it was reinforced in men's speech, while the older local form *ʔani* increasingly became restricted to women's speech.

Meanwhile, in the coastal dialects (region C), which had less contact with the dialects exclusively using *ʔana* (region A), a further case of analogical leveling took place: Women who used *ʔani* for 'I' also used *ḍarabni* to mean 'he hit me,' while men, using *ʔana* for 'I,' formed the equivalent phrase as *ḍarabna* (Behnstedt and Woidich 2010: 172). However, in region B, while speakers adopted the innovative form *ʔani* for the first-person singular feminine pronoun, they retained the older suffix pronoun *-ni* for both genders, under the influence of neighboring dialects in region A.

### Scenario 2: Cushitic

Looking westward across the Red Sea and southward along the opposite coast of the Gulf of Aden, we find several East Cushitic languages, such as Rendille, Oromo, and Sidaama, where the first-person singular pronoun for both genders is also *ani* (Appleyard 2011: 44). It is conceivable that the older Yemeni form *ʔani* was influenced by an ancient Cushitic substrate.<sup>16</sup> During the Arabization of the region, this older form *ʔani* might have initially coexisted with the newly introduced Arabic prestige form *ʔana* because of phonetic similarity.

Over time, the gender distinction between *ʔana* and *ʔani* may have emerged along the sociolinguistic patterns described above. In region C, the distinction was then analogically extended to suffix pronouns, while in region B, the common suffix pronoun was retained under the influence of adjacent region A.

---

<sup>15</sup> The areas east of the isogloss are substantially influenced by North Arabian dialects (Behnstedt and Woidich 2010: 182).

<sup>16</sup> Militarëv (1984) identified a Cushitic-like substrate in the Modern South Arabian languages (Blažek 2013: 2–3). The Cushitic-speaking population had later emigrated to Ethiopia or been assimilated by the Semitic-speaking population.

### Scenario 3: Ḥimyaritic

While traces of Cushitic influence in Yemeni dialects are difficult to pinpoint, remnants of pre-Arabic Semitic languages are more clearly identifiable (Watson 2018: 332–33). Like older Ḥimyaritic, some Yemeni dialects show perfect endings with *k* instead of the usual *t*: *katabku* instead of *katabtu* (Behnstedt and Woidich 2010: 156–157; Watson 2018: 321–322), cf. the form *qulk<sup>q</sup>lik* in section 3. A similar case might apply to the form *ʾani*. In fact, the form is attested in Ḥimyaritic (Robin 1991: 107):

(5) *anī Daybaḡa<sup>t</sup> bint Nawf ḡī Šaqr bin ḡī Murāṭīd ...*

‘I, Daybaḡa, daughter of Nawf of Šaqr, son of Murāṭīd ...’

(6) *anī Sam<sup>‘a</sup> bint ḡī Murāṭīd ...*

‘I, Sam<sup>‘a</sup>, daughter of Murāṭīd ...’

Since these attestations refer to women using *anī*, Behnstedt and Woidich (2010: 150) suggest that this form had already been grammaticalized as a feminine pronoun in Ḥimyaritic. However, this conclusion is premature, as no corresponding masculine pronoun with a different vocalization has been recorded. It is possible that *anī* was used for both genders (as in Hebrew).<sup>17</sup>

If we compare the geographic distribution of the *ʾana* vs. *ʾani* distinction with the remnants of Ḥimyaritic-speaking areas documented in the 10th century by Al-Hamdānī (cf. Figure 2), we find overlaps. He reported that Ḥimyaritic or Arabic mixed with Ḥimyaritic was still spoken in the highlands at that time (Behnstedt and Woidich 2010: 24).

While these areas lie somewhat further east than present-day dialect regions B and C, this discrepancy could be due to Al-Hamdānī’s focus on general linguistic evaluation rather than specific dialect features. If he primarily interacted with men, he may not have encountered the feminine *ʾani* form at all. Thus, the *ʾana* vs. *ʾani* split may have already existed in regions that he considered speaking ‘good Arabic’.

On the other hand, if Al-Hamdānī’s account is accurate, the gender distinction must have emerged over subsequent centuries—following the processes described for scenarios 1 and 2—spreading westward with further generalizations while being displaced in the east by dialects using *ʾana* (cf. Behnstedt and Woidich 2010: 150).

Due to the limited historical documentation of dialect variation in Yemen, it is difficult to determine the exact origins of the form *ʾani*. Among the three proposed hypotheses, Cushitic influence appears the least likely, as it assumes an exceptional degree of longevity and stability for the *ʾani* form. If there was indeed

<sup>17</sup> None of the other studies on Ḥimyaritic (Rabin 1951; Robin 1991; 2006) mention a gender distinction in the first-person singular pronoun.

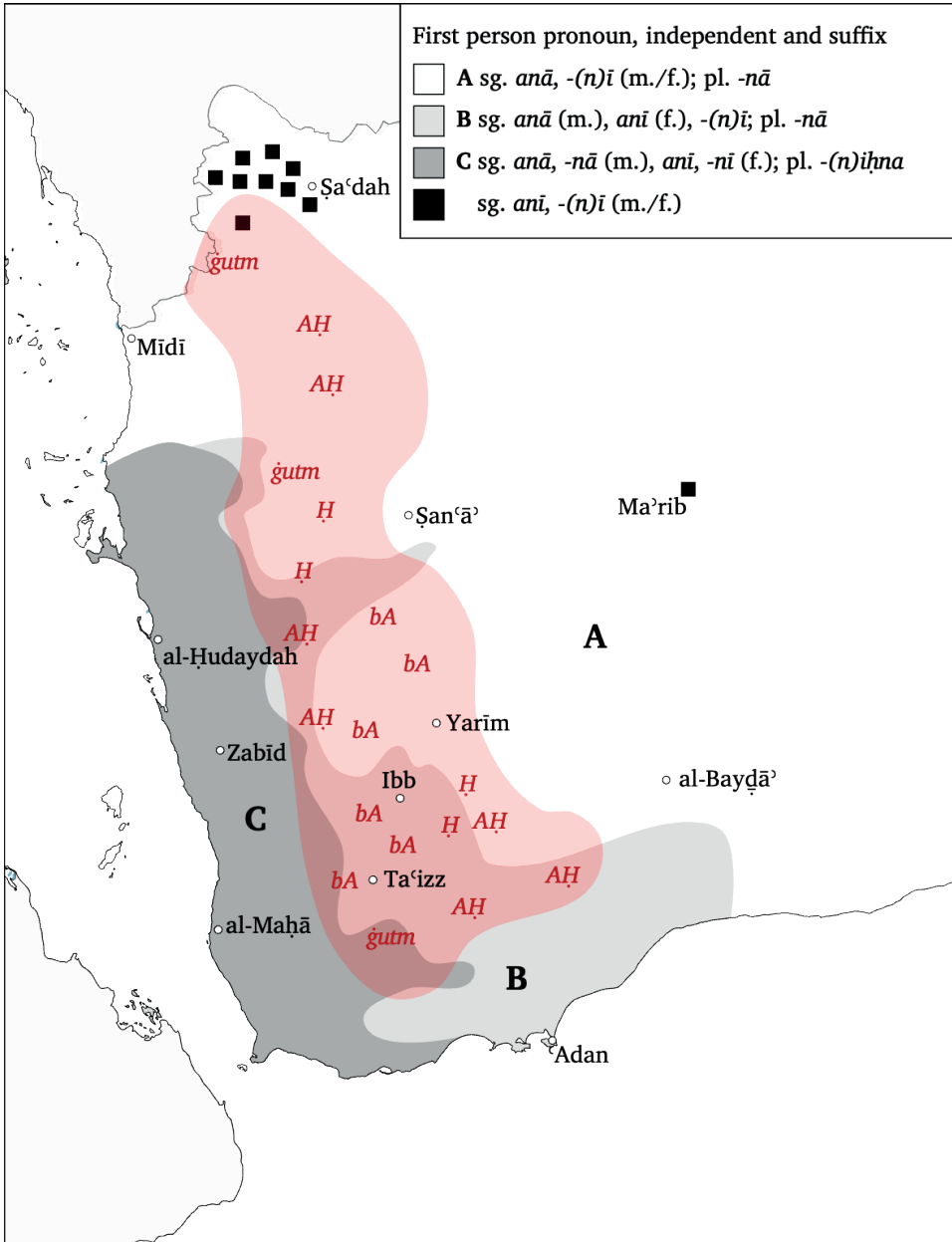


Figure 2. Distribution of Ḥimyaritic language remnants ( $\dot{H}$ ) and Ḥimyaritic-influenced Arabic ( $A\dot{H}$ ), bad Arabic ( $bA$ ) or unintelligible speech ( $\dot{g}utm$ ) in the 10th century according to Al-Ḥamdānī (10th c.), overlaid over Figure 1 (map based on Behnstedt and Woidich 2010: 157; Robin 1991: 83)

a Cushitic substrate, it would likely have left only faint traces in today's Arabic dialects and would almost certainly not have affected the pronominal system—one of the most resistant parts of a language to borrowing (Tadmor 2009: 67).

It remains unclear whether ʾani originated from a Bedouin dialect with ʾimāla, from Ḥimyaritic, or from an interaction between both influences. In any case, this raises a broader question: Why have distinctions between older forms and more recent dialectal forms, as well as between the forms of substrate languages (e.g., Aramaic, Coptic, Tamazight) and Arabic, not been similarly grammaticalized into gendered forms in other Arabic dialects?

## 4 Conclusion

While the phonological history of the two pronominal forms in Tocharian A remains unresolved, certain explanatory approaches align well with established sociolinguistic patterns. The collapse of case distinctions in pronouns led to the generalization of the accusative form in men's speech, while women's speech retained the older nominative form—at times even extending it hypercorrectively. Although regional contact influences that might have contributed to this distinction cannot be ruled out, there is no concrete evidence for them.

The gender-specific forms of the first-person singular pronoun in Yemeni Arabic dialects likely emerged from the competition between an older local form ʾani and a supralocal ʾana, in line with broader sociolinguistic trends observed in the Arab world. The precise origins of ʾani remain unclear—whether it stems from a Bedouin Arabic dialect, from pre-Islamic Ḥimyaritic, or from an interaction between the two. Nevertheless, the gender distinction in the first-person pronoun fits neatly into a broader pattern of proportional analogy with the second-person pronouns.

## Disclosure statement

No potential conflict of interest was reported by the author.

## ORCID

Kierán Meinhardt  <https://orcid.org/0009-0000-1250-807X>

## References

- Aikhenvald, Alexandra Y. 2000. *Classifiers: A Typology of Noun Categorization*. Oxford: Oxford University Press.
- Al-Wer, Enam. 2014. 'Language and Gender in the Middle East and North Africa'. In *The Handbook of Language, Gender, and Sexuality*, edited by Susan Ehrlich, Miriam Meyer-

- hoff, and Janet Holmes, 1st ed., 396–411. Chichester: Wiley Blackwell. <https://doi.org/10.1002/9781118584248.ch20>.
- Appleyard, David L. 2011. 'Semitic-Cushitic/Omotic Relations'. In *The Semitic Languages: An International Handbook*, edited by Stefan Weninger, 38–53. Handbücher Zur Sprach- Und Kommunikationswissenschaft 36. Berlin and Boston: De Gruyter Mouton.
- Bassiouney, Reem. 2020. *Arabic Sociolinguistics: Topics in Diglossia, Gender, Identity, and Politics*. 2nd ed. Washington, DC: Georgetown University Press.
- Behnstedt, Peter. 1985. *Die nordjemenitischen Dialekte: Atlas*. Vol. 1. Jemen-Studien 3. Wiesbaden: L. Reichert.
- Behnstedt, Peter. 2016. *Dialect Atlas of North Yemen and Adjacent Areas*. Handbook of Oriental Studies 114. Leiden and Boston: Brill.
- Behnstedt, Peter, and Manfred Woidich. 2010. *Arabische Dialektgeographie: Eine Einführung*. Handbuch der Orientalistik 78. Leiden and Boston: Brill.
- Berg, Lodewijk Willem Christiaan van den. 1886. *Le Ḥadhramout et les colonies Arabes dans l'archipel Indien*. Batavia: Imprimerie du gouvernement.
- Blažek, Václav. 2013. 'Levant and North Africa: Afroasiatic Linguistic History'. In *The Encyclopedia of Global Human Migration*, edited by Immanuel Ness, 1st ed. Chichester: Wiley Blackwell. <https://doi.org/10.1002/9781444351071.wbeghm815>.
- Brockelmann, Carl. 1908. *Grundriss der vergleichenden Grammatik der semitischen Sprachen*. Vol. 1. 2 vols. Berlin: Reuther & Reichard.
- Bross, Christoph. 2022. 'Untersuchungen zu den Personalpronomina im Tocharischen'. PhD thesis, München: Ludwig-Maximilians-Universität. [https://edoc.ub.uni-muenchen.de/30367/1/Bross\\_Christoph.pdf](https://edoc.ub.uni-muenchen.de/30367/1/Bross_Christoph.pdf).
- Carling, Gerd, Georges-Jean Pinault, and Melanie Malzahn. 2014. 'A 66'. CEToM: A Comprehensive Edition of Tocharian Manuscripts. 31 May 2014. <https://cetom.univie.ac.at/?m-a66>.
- Connors, Thomas J. 2008. 'Tengger Javanese'. PhD thesis, New Haven: Yale University.
- Corbett, Greville. 1991. *Gender*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- Costa, Januacele Francisca da. 1999. 'Ya:thê, a última língua nativa no nordeste do Brasil. Aspectos morfo-fonológicos e morfo-sintáticos'. PhD thesis, Recife: Universidade federal de Pernambuco.
- Cysouw, Michael. 2003. *The Paradigmatic Structure of Person Marking*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.
- Dickinson, Connie. 2002. 'Complex Predicates in Tsafiki'. PhD thesis, Eugene: University of Oregon.
- Diem, Werner. 1973. *Skizzen jemenitischer Dialekte*. Beirut Texts and Studies 13. Beirut: Franz Steiner.
- Dores de Oliveira, Maria das. 2006. 'Ofayé, a língua do povo do mel fonologia E gramática'. PhD thesis, Maceió: Universidade federal de Alagoas. [https://dlc.library.columbia.edu/catalog/ldpd:504833/bytestreams/content/content?filename= Maria + das + Dores + de + Oliveira.pdf](https://dlc.library.columbia.edu/catalog/ldpd:504833/bytestreams/content/content?filename=Maria+das+Dores+de+Oliveira.pdf).
- Fortson, Benjamin W. 2010. 'Tocharian'. In *Indo-European Language and Culture: An Introduction*, Benjamin W. Fortson, 2nd ed., 400–413. Blackwell Textbooks in Linguistics 19. Malden, MA: Wiley-Blackwell.
- Hachimi, Atiqa. 2006. 'Gender'. In *Encyclopedia of Arabic Language and Linguistics*, edited by Kees Versteegh, Mushira Eid, Alaa Elgibali, Manfred Woidich, and Andrzej Zaboriski, 2:155–164. Leiden and Boston: Brill.

- Hermann, Eduard. 1922. 'Königlich preußische Turfanexpeditionen. Tocharische Sprachreste, I. Band. Die Texte: A. Transkription'. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete der Indogermanischen Sprachen* 50 (3/4): 296–314.
- Hinds, John. 1986. *Japanese: Descriptive Grammar*. Croom Helm Descriptive Grammars. London and New York: Taylor and Francis.
- Isaksson, Bo. 1999. 'The Non-Standard First Person Singular Pronoun in the Modern Arabic Dialects'. *Zeitschrift für Arabische Linguistik* 37: 54–83.
- Ivanov, Vjačeslav Vsevolodovič. 2008. 'Tibetskie kal'ki v toxarskix tekstax'. In *Trudy po étimologii indoevropskix i drevneperedneaziatskix jazykov*, 2:588–295. Moskva: Znak.
- Jasanoff, Jay H. 1989. 'Language and Gender in the Tarim Basin: The Tocharian 1 Sg. Pronoun'. *Tocharian and Indo-European Studies* 3: 125–418.
- Jesus Silva, Léia de. 2011. 'Morphosyntaxe du Rikbaktsa (Amazonie brésilienne)'. PhD thesis, Paris: Université Denis Diderot – Paris 7. [https://etnolinguistica.wdfiles.com/local--files/tese%3Asilva-2011b/silva\\_2011\\_rikbaktsa.pdf](https://etnolinguistica.wdfiles.com/local--files/tese%3Asilva-2011b/silva_2011_rikbaktsa.pdf).
- Kibebe Tsehay Taye. 2015. 'Documentation and Grammatical Description of Chabu'. PhD thesis, Addis Ababa: Addis Ababa University.
- Krause, Wolfgang, and Werner Thomas. 1960. *Tocharisches Elementarbuch: Grammatik*. Heidelberg: Universitätsverlag Winter.
- Labov, William. 2006. *Principles of Linguistic Change: Social Factors*. Vol. 2. Language in Society 29. Malden, MA: Blackwell.
- Lerthirunwong, Malai. 1980. 'Syntactic Description of So: An Austroasiatic Language in Thailand'. Master's thesis, Mahidol University. <http://archives.sealang.net/mahidol>.
- Lionnet, Florian. 2010. 'Laal: An Unclassified Language of Southern Chad'. Paper presented at the Workshop on Language Isolates in Africa, Lyon, 3–4 December 2010.
- Militarëv, Aleksandr. 1984. 'Sovremennoe sravnitel'no-istoričeskoe afrazijskoe jazkoznanie: čto ono možet dat' istoričeskoj nauke?'. In *Lingvističeskaja rekonstrukcija i drevnejšaja istorija Vostoka. Tezisy i doklady konferencii*, edited by Igor Vardul', 3: 3–26. Moskva: Nauka.
- Plank, Frans, and Wolfgang Schellinger. 1997. 'The Uneven Distribution of Genders over Numbers: Greenberg Nos. 37 and 45'. *Linguistic Typology* 1: 53–101.
- Rabin, Chaim. 1951. *Ancient West Arabian*. London: Taylor's Foreign Press.
- Robin, Christian Julien. 1991. *L'Arabie antique de Karib'ıl à Mahomet : nouvelles données sur l'histoire des Arabes grâce aux inscriptions*. Aix-en-Provence: Édisud.
- Robin, Christian Julien. 2006. 'Ĥimyaritic'. In *Encyclopedia of Arabic Language and Linguistics*, edited by Kees Versteegh, Mushira Eid, Alaa Elgibali, Manfred Woidich, and Andrzej Zaborski, 2:256–261. Leiden and Boston: Brill.
- Rosenhouse, Judith. 2006. 'Bedouin Arabic'. In *Encyclopedia of Arabic Language and Linguistics*, edited by Kees Versteegh, Mushira Eid, Alaa Elgibali, Manfred Woidich, and Andrzej Zaborski, 1:259–269. Leiden and Boston: Brill.
- Sadiqi, Fatima. 2011. 'Language and Gender'. In *Encyclopedia of Arabic Language and Linguistics Online*, edited by Lutz Edzard and Rudolf de Jong. Accessed 20 October 2025. [http://dx.doi.org/10.1163/1570-6699\\_eall\\_EALL\\_COM\\_vol2\\_0083](http://dx.doi.org/10.1163/1570-6699_eall_EALL_COM_vol2_0083).
- Sands, Bonny. 2013. 'Hadza'. In *The Khoesan Languages*, edited by Rainer Vossen, 38–42, 89–90, 107–24, 265–74. Routledge Language Family Series. London and New York: Routledge.
- Shachmon, Ori. 2022. *Temonit: The Jewish Varieties of Yemeni Arabic*. Semitica Viva 62. Wiesbaden: Harrassowitz.

- Skirgård, Hedvig, Hannah J. Haynie, Harald Hammarström, Damián E. Blasi, Jeremy Collins, Jay Latache, Jakob Lesage, et al. 2023. 'Grambank V1.0'. Zenodo. <https://doi.org/10.5281/ZENODO.7740140>.
- Smyth, David. 2002. *Thai: An Essential Grammar*. Essential Grammars. London: Routledge.
- Tadmor, Uri. 2009. 'Loanwords in the World's Languages: Findings and Results'. In *Loanwords in the World's Languages*, edited by Martin Haspelmath and Uri Tadmor, 55–75. Berlin and Boston: Walter de Gruyter. <https://doi.org/10.1515/9783110218442.55>.
- Takács, Gábor. 2011. 'Semitic-Egyptian Relations'. In *The Semitic Languages: An International Handbook*, edited by Stefan Weninger, 7–18. Handbücher Zur Sprach- Und Kommunikationswissenschaft 36. Berlin and Boston: De Gruyter Mouton.
- Trudgill, Peter. 2000. *Sociolinguistics: An Introduction to Language and Society*. 4th ed. London: Penguin Books.
- Vallejos Yopán, Rosa. 2010. 'A Grammar of Kokama-Kokamilla'. PhD thesis, Eugene: University of Oregon. [https://etnolinguistica.wdfiles.com/local--files/tese%3Avallejos-2010/vallejos\\_2010\\_kokama.pdf](https://etnolinguistica.wdfiles.com/local--files/tese%3Avallejos-2010/vallejos_2010_kokama.pdf).
- Vicente, Ángeles. 2021. 'Statistical Gender Indexicality in the Grammar of Vernacular Arabic Varieties: Types and Locus'. *Brill's Journal of Afroasiatic Languages and Linguistics* 13 (2): 217–238. <https://doi.org/10.1163/18776930-01302001>.
- Watson, Janet. 2018. 'South Arabian and Arabic Dialects'. In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*, edited by Clive Holes, 316–334. Oxford Studies in Diachronic and Historical Linguistics. Oxford: Oxford University Press.
- Werbeck, Wolfgang. 2001. *Laut- und Formenlehre des nordjemenitisch-arabischen Dialekts von Manāḥa*. Arabica Rhema. Münster: Rhema.