The representation of gender on local person pronouns

Zorica Puškar-Gallien (ZAS Berlin)

The issue: Despite gender being an extensively studied topic, what has evaded closer scrutiny of the previous literature are languages in which local person (i.e. 1^{st} and 2^{nd} person) pronouns show gender distinctions (see e.g. Audring 2008; Siewierska 2013*a*). I present a survey of languages with morphological gender distinctions on local person, arguing that gender as a feature must be encoded higher in the structure than person and number, and as such it can affect morphological processes such as allomorphy. **Empirical Focus:** The World Atlas of Language Structures lists 19 languages with gender distinctions in 3^{rd} person and 1^{st} and/or 2^{nd} person and 2 languages with gender distinctions in 1^{st} or 2^{nd} person but not 3^{rd} (Siewierska 2013*b*). Extending the inquiry to the families that the languages listed belong to, I have so far identified 54 languages with gender on local person, belonging to 18 families, and 2 isolates. Table 1 presents a summary of the patterns found thus far.

sg	3 + 2 + 1	Korana, Hadza, Ngala
sg	3 + 2	Abkhaz, Berber (Riffian), Amharic, Arabic, Aramaic, Hebrew, Musey, Kera, Lele, Hausa, Mupun, Kulere, Fyer, Ron (Bokkos, Dafo-Butura), Zari, Miya, Beja, Coptic, Nama, Tunica
sg	3+ 1	Cocama-Cocamilla
sg	2 + 1	Andi, Paez
sg	2	Berber (Tamasheq), Sha, Kofyar, Burungue, Iraqw, Moken (Dung, Rawai, Surin), Minangkabau, Cholon
pl	3 + 2 + 1	Korana, Hadza, Spanish, Provencal, Slovenian, Berber (Tamasheq, Riffian), Gagadu, Nama
pl	3 + 2	Arabic, Hebrew, Beja
pl	2	Andi
du	3 + 2 + 1	Korana, Lithuanian, Dumo, Bora, Murui Huitoto, Slovenian, Nama, Djeebbana, Murrinhpatha, Burarra, Nunggunuyu, Anindilyakwa, Gagadu, Touo
du	3 + 2	Kamoro, Tunica

Table 1: Languages with gender on local person pronouns

The data permit several initial generalizations. Gender on all three persons is present in many languages of the sample, across all three number values. Some languages have gender distinctions on 1^{st} and 2^{nd} person to the exclusion of 3^{rd} (Andi, Paez). Such a constellation is present only in the singular. Gender distinctions only on 2^{nd} person are present on 10 languages in the SG and only one in the PL (Andi), none in the DU. Otherwise, gender distinctions in 3^{rd}+2^{nd} person are present in several languages across all three number values. It may also be noted that languages where gender is present on 2^{nd}+3^{rd} person in the PL show the same pattern in the singular. The classification in Table 1 exposes gaps in the following conditions: (i) only 1^{st} person across all numbers, (ii) 2^{nd}+1^{st} person in plural and dual, and (iii) 3^{rd}+1^{st} person across all numbers. In sum, this allows for more formal generalizations as in (1)–(2).

(1) Having gender distinctions in 1^{st} person singular entails having gender distinctions in 2^{nd} person, or both in 2^{nd} and 3^{rd} person.

(2) Having gender distinctions in 1^{st} person non-singular entails having gender distinctions on 2^{nd} and 3^{rd} person.

Table 2 represents the personal pronouns in Korana (Khoe-Kwadi, South Africa). The paradigm is noteworthy as the most complete one in the sample collected by Siewierska (2013*b*). It distinguishes between M, F and Common gender in both inclusive and exclusive 1^{st} person contexts, as well as in

other persons. Given the generalizations in (1)-(2) and paradigms such as the one in Table 2, two main questions arise: (Q1) How does feature entailment get accounted for? (Q2) What does the morphological structure (e.g. morpheme ordering) tell us about the internal structure of a pronoun?

(Towards a) proposal: ϕ-features have complex internal structure in the form of hierarchically organised sub-features (Harley and Ritter 2002). Person features can be further decomposed such that the complexity of representation increases from the 3^{rd} towards the 1^{st} person (McGinnis 2005; Georgi 2012, 2013; Nevins 2007; Béjar and Řezáč 2009; Preminger 2014; Deal 2015; Kalin 2019). I take 3^{rd} person to be represented by the general node π, 2^{nd} person includes an additional node [Addressee [π]] and 1^{st} person includes an additional node [Speaker [Addressee [π]]]. The entailment relation between person features alone partially accounts for the gaps in (1)-(2); having 1^{st} person is impossible without including the feature [Addressee]. This indicates that whatever is responsible for encoding gender on 2^{nd} person will be included in the formal description of 1^{st} person, but not the other way round (making 2^{nd} person more independent). The root node π can also be absent, resulting in absence of person features. I will adopt Harley and Ritter (2002) representation of number involving subfeatures [Group] and [Minimal], whose combinations may derive SG, PL and DU, where SG can also be the absence of number. I follow Hammerly 2018; Puškar 2018; Foley and Toosarvandani 2022; Caha 2021; Adamson and Anagnostopoulou 2024 in assuming that gender can be represented by a combination of features [CLass], [F], [ANIMate], [HUMan] (hierarchical entailment relations of number and gender omitted for reasons of space). Person and number features have been argued to reside on two separate projections, person being lower than number (Moskal 2015; Harbour 2016; Smith et al. 2018; van Urk 2018). E.g. if pronouns are decomposed into person number and case, number comes in between person (pronominal base) and case (Noyer 1992; Trommer 2002; Harbour 2007, 2008, 2016; Arregi and Nevins 2012). The Mirror Principle (Baker 1985; Brody 2000; Brody and Szabolcsi 2003) indicates a lower base position of person with respect to number (3). I argue that grammatical gender heads its own projection CLP above the #P.

		SG	DU	PL
1	C		*sa-m*	*sa-da*
INCL	F		*sa-sam*	*sa-se*
	M		*sa-kham*	*sa-tje*
1	C		*s-im*	*si-da*
EXCL	F	*ti-ta*	*si-sam*	*si-se*
	M	*ti-re*	*si-kham*	*si-tje*
2	C		*sa-khaoo*	*sa-du*
	F	*sa-s*	*sa-saro*	*sa-sao*
	M	*sa-ts*	*sa-kharo*	*sa-kao*
3	C	*ll'di-'i*	*ll'di-kha*	*ll'dine*
	F	*ll'di-s*	*ll'di-sara*	*ll'dide*
	M	*ll'di-b*	*ll'di-khara*	*ll'dku*

Table 2: Korana (Siewierska 2013*b*)

(3) $[_{\text{CLP}}\ \text{CL}_{[\text{HU, ANIM, F}]}\ [_{\#\text{P}}\ \#_{\text{MIN, GROUP}}\ [_{\pi}\ \pi_{\text{ADDRESSEE, SPEAKER}}\ ]]]$

This predicts that given the presence of all three ϕ-features on a pronoun, they will align according to the hierarchy π>#>CL (c.f Noyer 1992). Some languages in the sample conform to this prediction (Andi, Arabic, Berber, Bora, Djeebbana, Gagadu, Nama, Provencal, Spanish, Lithuanian, Slovenian, Korana), while others challenge it by showing a gender-related change in the stem (Abkhaz, Hadza, Iraqw, Miya, Musey, Ron (Fyer, Bokkos, Daffo-Butura), Ngala, Kamoro, Cholon, Cocama-Cocamilla), stem suppletion (Kera, Mokken (Dung, Rawai, Surin), Mupun), or are not easily segmentable. I argue that while the underlying structure is the same, the superficial differences are due to language-specific ways of exponing the individual features. For those that challenge this proposal, I will argue that CLP is a locality domain contributing to the morphological realisation of the pronominal base, such that it can trigger suppletion, which can be analysed as contextual allomorphy in the sense of Moskal (2015); Smith et al. (2018); McFadden (2018).