A structural constraint on typological variation in personal pronouns*

Silvia Terenghi

Utrecht University

1. Introduction

Building on a typologically varied sample of independent pronominal paradigms, this paper shows a hitherto unnoticed generalisation with respect to the variation in the featural make-up of personal pronouns: namely, for a pronominal paradigm to encode (any type of) number oppositions, that paradigm needs to make at least a ternary person distinction (1–2–3). This generalisation is shown to fall from the internal structure of pronominal forms and, more specifically, from the structural relations among the relevant features. Concretely, an implicational relation can be defined across the active features, whereby less deeply embedded features may not be present if more deeply embedded ones are absent. In turn, this account is rooted in the action-on-lattice nature of person and number features.

This paper is organised as follows. Section 2 lays out the assumptions on which the argumentation is based, both concerning person and number features and regarding the (lower) internal structure of personal pronouns. Section 3 swiftly illustrates the sample on which this work is based, while Section 4 introduces the aforementioned typological generalisation and further formalises it. Section 5, instead, puts forth an account for this generalisation in terms of implicational relations across features. Section 6 concludes.

2. Assumptions: Action-on-lattice features

This paper is rooted into Harbour's (2014a, 2016) action-on-lattice approach to person and number features. Under this approach, person and number features are construed as denoting lattices (that is, power sets of (subsets of) the ontology). These, by means of their values (+ or −), perform set-theoretic operations on their complements, which are in turn taken to denote (larger) lattices. These operations result in the partition of the complement lattice into smaller subsets, which correspond to the different person and number cate-

*I would like to thank audiences at NELS 53 and CGG 31, as well as Roberta D'Alessandro and Martin Everaert, for discussion and feedback. This project was partly funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement: CoG 681959_MicroContact).

gories, according to which feature(s) were active and which operation(s) were performed. The relevant features, as well as their distribution within the internal structure of personal pronouns, are presented in what follows.

2.1 Person

Following Harbour (2016), this paper assumes two person features, [author] and [participant], which denote the power sets of different subsets of the person ontology (the latter consists of three basic atoms: speaker *i*, hearer *u*, and other(s), *o*), as follows:

(1) a. $[\![\text{author}]\!] = \{i\}$ (Harbour 2016:73–74)
b. $[\![\text{participant}]\!] = \{i, iu, u\}$ (Harbour 2016:73–74)

[Author] and [participant] (successively) perform set-theoretic operations on their complement, the π head (the power set of the improper subset of the person ontology; see (2)) or the result of a previous operation on the π head, thanks to their values ($+$, which induces disjoint addition; and $-$, which induces joint subtraction).

(2) $[\![\pi]\!] = \{i_o, iu_o, u_o, o_o\}$[1] (Harbour 2016:73–74)

The different operations performed on π partition this set into different subsets, yielding contrastive person categories and, in turn, different person systems: their different featural derivations are summarised in (3):

(3) *Person systems*

<table>
<tr><th></th><th>Unary
—</th><th>$\text{Binary}_{[\text{pt}]}$
1/2|3</th><th>$\text{Binary}_{[\text{au}]}$
1|2/3</th><th>Ternary
1|2|3</th><th>Quaternary
1E|1I|2|3</th></tr>
<tr><td>i_o</td><td rowspan="4">π</td><td rowspan="3">$+\text{pt}(\pi)$</td><td rowspan="2">$+\text{au}(\pi)$</td><td rowspan="2">$+\text{pt}(+\text{au}(\pi))$</td><td>$+\text{au}(-\text{pt}(\pi))$</td></tr>
<tr><td>iu_o</td><td>$+\text{au}(+\text{pt}(\pi))$</td></tr>
<tr><td>u_o</td><td rowspan="2">$-\text{au}(\pi)$</td><td>$+\text{pt}(-\text{au}(\pi))$</td><td>$-\text{au}(+\text{pt}(\pi))$</td></tr>
<tr><td>o_o</td><td>$-\text{pt}(\pi)$</td><td>$-\text{pt}(\pm\text{au}(\pi))$</td><td>$-\text{au}(-\text{pt}(\pi))$</td></tr>
</table>

2.2 Number

This paper is only concerned with exact numbers and adopts the exact number feature system put forth by Harbour (2014a). This system only includes two number features, [atomic] and [minimal]:

(4) a. $[\pm\text{atomic}] = \lambda x\ (\neg)\text{atom}(x)$ (Harbour 2014a:202)
b. $[\pm\text{minimal}] = \lambda \text{P}\ \lambda x\ (\neg)\neg\exists\ y\ (\text{P}(y) \wedge y \sqsubset x)$ (Harbour 2014a:202)

[1] Subscript *o* indicates in short that any number of ‘others’ may be present, including zero.

The two number features (successively) operate on the lattice regions denoted by their complements and partition them into regions containing or not containing atoms and into regions containing or not containing minimal elements, respectively. As such, different number categories are yielded, which constitute different number systems; their featural derivations are given in (5):

(5) *Number systems*

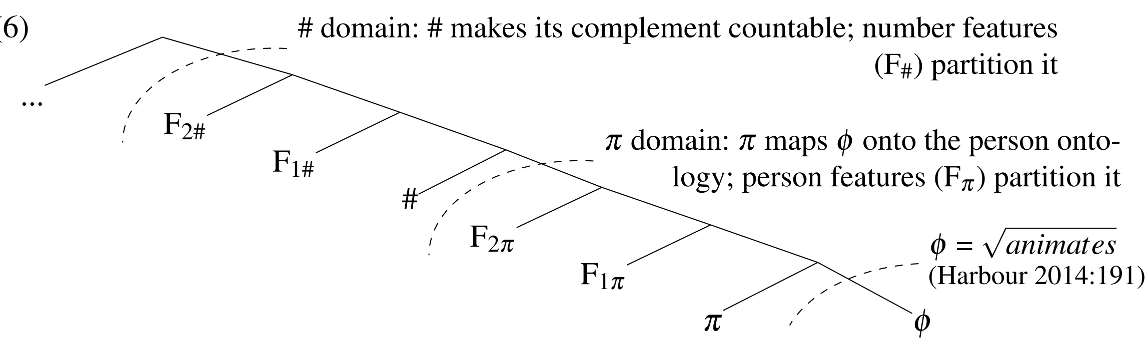
<table>
<tr><th></th><th>0
—
—</th><th>$2_{[at]/[min]}$
SG|PL
MIN|AUGM</th><th>$3_{[at]/[min]}$
SG|DU|PL
MIN|U–A|AUGM</th><th>$+_{[at]/[min]}$
SG|DU|...|PL
MIN|U–A|...|AUGM</th></tr>
<tr><td>{*a*}</td><td rowspan="4">P</td><td>+at/min(P)</td><td>+min(+at/min(P))</td><td>...(+min(+at/min(P)))</td></tr>
<tr><td>{*b*, *c*}</td><td rowspan="3">−at/min(P)</td><td>+min(−at/min(P))</td><td>...(+min(−at/min(P)))</td></tr>
<tr><td>{*d*, *e*, *f*}</td><td rowspan="2">−min(−at/min(P))</td><td>...(−min(−at/min(P)))</td></tr>
<tr><td>{*g*, ...}</td><td>...(−min(−at/min(P)))</td></tr>
</table>

2.3 Number above person

Finally, a 1 Feature–1 Head architecture for syntax is assumed in this paper. This is a stark departure from the accounts proposed by Harbour and revised in the foregoing, but is fully compatible with the action-on-lattice framework, once the Axiom of Extension is assumed (see also remarks by Harbour 2014a:192 in this sense). Moreover, number is taken to be is structurally higher than person, following morphological and semantic arguments advanced, among others, by Harbour (2016), Vanden Wyngaerd (2018), and Ackema and Neeleman (2018).

Given these two structural assumptions, for the purposes of this paper the (low) internal structure for personal pronouns is assumed to be as follows:[2]

(6)



In what follows, the structure in (6) is shortened by only referring to the relevant sequence of features/functions: $(F_{2\#}(F_{1\#}(F_{2\pi}(F_{1\pi}(...)))))$. Moreover, different pronominal

[2]This structure further minimally composes with a DP-layer, which is inconsequential here and will not be dealt with; this work also abstracts away from gender features. For a more extended discussion on both these issues and the other assumptions reviewed above, see Terenghi 2023:Appendix D).

paradigms are indicated by combining shorthands for the person domain (7a) and shorthands for the number domain (7b):

(7) a. $B_{[au]}$: Binary system, [author]
$B_{[pt]}$: Binary system, [participant]
T: Ternary system
Q: Quaternary system

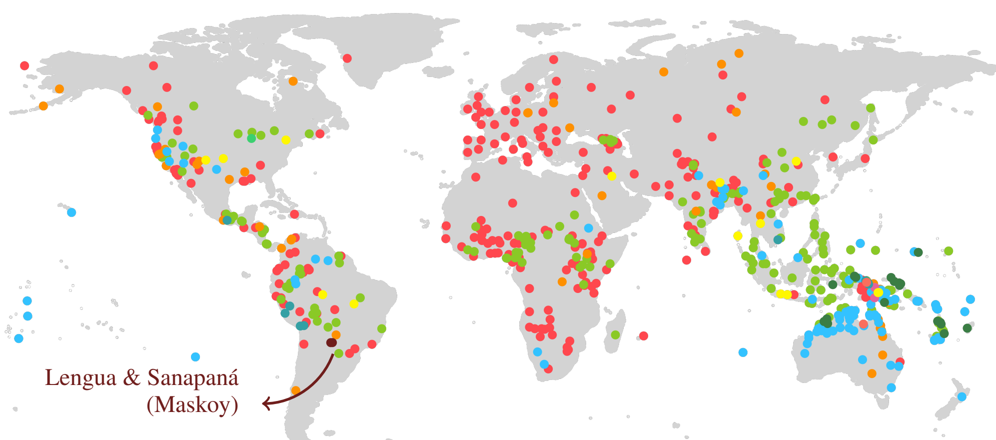
b. 0: no number opposition
2: binary number opposition
3: ternary number opposition
+: larger number opposition

For instance, T_2 indicates a ternary person paradigm (1|2|3) which makes a two-way number distinction (SG|PL or MIN|AUGM).

3. The sample

The present work focuses on independent pronominal paradigms, as these are generally acknowledged to display more paradigmatic oppositions than dependent ones (Siewierska 2004:112–113; Cysouw 2009:311–315). The pronominal paradigms examined here have been collected from the following sources: Forchheimer 1953; Noyer 1992; Corbett 2000; Harley and Ritter 2002; Siewierska 2004; Bhat 2004; Cysouw 2009; Smith 2011; Harbour 2016. The resulting sample includes 673 languages (125 families; 234 *genera*; classification based on Dryer 2013); their distribution is shown in (8):[3]

(8)



An full overview of the paradigms attested in this sample is given in (9), on the next page.

4. Formalisation

A pattern that emerges from (9) concerns the compatibility of person and number systems: the activation of both person features is compatible with any number system (column '$F_{1\pi}$, $F_{2\pi}$'); instead, the activation of one person feature seems to only be possible in a number-neutral system (column '$F_{1\pi}$'). Conversely, (9) shows that, if a paradigm encodes number

[3]For the full sample, its classification, and its genealogical classification, see Terenghi 2023:Appendix D.

(9) *Attested paradigms: A quantitative overview*

	$F_{1\pi}$, $F_{2\pi}$		$F_{1\pi}$	
$F_{1\#}$, $F_{2\#}$, ...	● Q_+: ...(±min(±at/min(±au(±pt(...)))))	n=19	—	
	● T_+: ...(±min(±at(±pt(±au(...)))))	n=2	—	
$F_{1\#}$, $F_{2\#}$	● Q_3: ±min(±at/min(±au(±pt(...))))	n=102	—	
	● T_3: ±min(±at(±pt(±au(...))))	n=55	—	
$F_{1\#}$	● Q_2: ±at/min(±au(±pt(...)))	n=197	—	
	● T_2: ±at(±pt(±au(...)))	n=267	● B: ±at(±au(...))	n=2
ø	● Q_0: ±au(±pt(...))	n=10	● B_{pt}: ±pt(...)	n=1
	● T_0: ±pt(±au(...))	n=16	● B_{au}: ±au(...)	n=2

oppositions, it typically also displays a full set of person oppositions (that is, it is a Q or a T system, derived by two features). A seeming exception to this last observation is provided by B systems, i.e., cognate Lengua and Sanapaná (Harbour 2016:55–56):[4]

(10) *Sanapaná (Mascoian; Harbour 2014b:127)*

	SG	PL
1	ko'o	enenko'o
2/3M	hlejap	hlengap
2/3F	hleja	hlenga

As shown by (10), Sanapaná opposes first person and non-first person forms, derived by means of [±author], but it also encodes a two-way number opposition, derived by means of [±atomic]. At face-value, this is in plain contradiction with the observation laid out above, as the Sanapaná system displays number oppositions despite not presenting a full set of person features. However, upon closer inspection, it can be argued that at least Sanapaná is in fact more akin to a T_2 system, but with a 2–3 syncretism. The main consideration that points in this direction concerns the contrastive encoding of 3rd person in the syntax. Contrary to the view that the language instantiates a binary person syntax, in fact, it can be observed that Sanapaná does contrastively encode 3rd person syntax for inanimate referents: this is achieved by means of demonstrative forms, which further consistently display a 3rd person syntax (Gomes 2013:228–231). This is taken here to indicate that the syntax does have the means to distinguish between 2nd and 3rd person and that the two simply happen to be realised by the same form in the case of animate referents: as such, Sanapaná would display syncretism between 2nd and 3rd person, rather than completely conflating of two categories into one (the "non-author").

[4]The discussion in what follows exclusively focuses on Sanapaná due to the lack of accessible information on Lengua.

Given the foregoing discussion, this paper upholds the descriptive generalisation whereby number features can be regarded as “parasitic” on person features; in fact, it makes a stronger claim, namely that the paradigms not attested in this sample are in fact banned:

(11) *Pronominal paradigms: A formal generalisation on typological variation*

	$F_{1\pi}$, $F_{2\pi}$	$F_{1\pi}$
$F_{1\#}$, $F_{2\#}$, ...	Q_+ & T_+	*
$F_{1\#}$, $F_{2\#}$	Q_3 & T_3	*
$F_{1\#}$	Q_2 & T_2	B*
ø	Q_0 & T_0	B_{pt} & B_{au}

Said otherwise, only pronominal paradigms that encode a full set of person features, thus yielding a ternary or a quaternary person-based opposition, may encode any number distinctions; and, conversely, if a pronominal paradigm lacks one (or more) person features, that paradigm must also be number-neutral:

(12) a. $...(F_{1\#}(F_{2\pi}(F_{1\pi}(...)))))$
b. $*...(F_{1\#}(F_{1\pi}(...)))$

Note that this generalisation partly contradicts Greenberg’s (1963) Universal 42 (“All languages have pronominal categories involving at least three persons and two numbers”, Greenberg 1963:75; see also Tvica’s (2017) Person–Number Universal), in that it explicitly allows for pronominal systems with less than two numbers, besides making no claims about (and, in fact, being fully compatible with) the availability of less than three persons.

5. A first-pass structural account

This section proposes a preliminary analysis for the generalisation that emerged from the pronominal data, as detailed in the foregoing. Before doing so, it is necessary to rephrase the generalisation in (12) in fully structural terms; assuming a 1 Feature–1 Head architecture for syntax (see again Section 2.3), (12) amounts to stating that less deeply embedded features cannot be merged if more deeply embedded ones are not merged; and, conversely, if one (more deeply embedded) person feature is missing, no (structurally higher) number feature may be present. In other words, an implicational relation holds within the sequence of heads in the internal structure of pronouns.

This is immediately reminiscent of approaches to language acquisition and structure building such as the truncation model (Rizzi 1993/4) and the Growing Trees view (Friedmann et al. 2021). Under these models, acquisition is argued to closely adhere to the structure of the tree moving bottom-up and banning “gaps” in the structure: no two structurally discontinuous portions of the tree may be acquired, to the exclusion of a third portion of the tree which is structurally between them. Here, this implicational model is extended from the clausal domain, for which it was originally proposed, to the functional sequence at word-internal level.

Preliminary evidence for the implicational relation that holds of the sequence of heads internal to personal pronouns can be sought in language acquisition. Despite the unavailability to date of a comprehensive investigation of the ordering of acquisition of person categories, in fact, foundational studies in this respect (Hanson 2000; Harley and Ritter 2002) report that pronominal paradigms start out in the singular; once the full set of person distinction has been acquired, children acquire the plural forms. Importantly, at the first stage of acquisition, singular is not contrastive in the system, as so plural forms are available; therefore, it can be more generally said that number distinction are acquired *after* person distinctions.

While so far this proposal is strictly structural in nature and depends on the assumptions in this respect discussed around (6), it can be speculated that the fact that number features are parasitic on the full set of person features being merged in the functional sequence is ultimately to be traced back to the action-on-lattice nature of the relevant features. As explained in Section 2, action-on-lattice features denote sets that perform set-theoretic operations on their complement. As such, and for functional application to be successful, the "internal" structure of the arguments of the relevant functions must be suitable.

Let us consider in this respect the two bipartitions: the speaker-based bipartition includes one joint non-speaker category (13a), whereas the participant-based one encompasses a joint participant category (13b):

(13) a. Minimal non-speaker category: $\{u, o\}$
b. Minimal participant category: $\{i, u\}$

In these configurations, two discourse atoms are minimally present at once: the hearer and the other(s) are simultaneously present in (13a); the speaker and the hearer are simultaneously present in (13b). This state of affairs can be seen as fundamentally incompatible with the application of [+atomic] and [+minimal]: the complement of these features needs in fact to have an atomic structure (or at least a minimal unit) for the operation to apply. Thus, it can be tentatively suggested that the application of number features is blocked by the unavailability of suitable arguments. Ultimately, then, the logical structure which underlies the definition of action-on-lattice features sets basic, logical constraints on how the functional sequence is assembled.

Before concluding, it is worth noting that the implicational formalisation of the structural relation does not rely on any *ad hoc* mechanism; as such, the ban on gaps in the functional sequence should be conceived as a general constraint that holds of any sequence of action-on-lattice heads, and not one that is only relevant in the derivation of the contingency of number features on person features explored in this work. Independent evidence that the same implicational relation holds of the functional sequence more widely is provided by diachrony, as I argued in Terenghi 2023. More specifically, person and number systems derived by two features can be observed to be diachronically unstable, yielding simpler paradigms derived by one single feature. Whenever this is the case. feature loss can be shown to proceed from the outside in. In the number domain, this amounts, among others, to the well-documented loss of the dual semantics, driven by the loss of $F_{2\#}$ (that is: [±minimal]). Likewise, the last person feature to be merged in the functional spine of demonstrative elements ($F_{2\pi}$: [±participant] in ternary systems and [±author] in quaternary systems) may undergo loss, yielding binary demonstrative systems; however, this is only possible if no number feature is merged, following the structural constraint identified in this work.

6. Conclusions

This paper identified a novel generalisation holding on independent pronominal paradigms: for a pronominal paradigm to encode (any type of) number oppositions, that paradigm needs to make at least a ternary person distinction (1–2–3). This typological description was first restated in featural terms, where it amounts to stating that number features may only be merged in the internal structure of personal pronouns if the full set of person features (namely: two) is also merged. In turn, building on the assumption of a 1 Feature–1 Head architecture for syntax, the featural generalisation was reduced to a structural constraint, whereby less deeply embedded heads (i.e., number features) may only be merged if the sequence of more deeply embedded heads (i.e., person features) is present, banning gaps within the functional sequence. As such, it was concluded that the (un)attested patterns of typological variation in independent pronominal paradigms are structurally constrained. Finally, this constraint was preliminarily argued to be rooted in the action-on-lattice nature of person and number features, and specifically in the restrictions that number features pose on the structure of their complement set.

References

Ackema, Peter, and Ad Neeleman. 2018. *Features of Person: From the inventory of persons to their morphological realization*. Cambridge, MA: The MIT Press.

Bhat, Darbhe Narayana Shankara. 2004. *Pronouns*. Oxford/New York: Oxford University Press.

Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.

Cysouw, Michael. 2009. *The Paradigmatic Structure of Person Marking*. Oxford/New York: Oxford University Press.

Dryer, Matthew S. 2013. Genealogical Language List. In *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. URL `https://wals.info/languoid/genealogy`.

Forchheimer, Paul. 1953. *The Category of Person in Language*. Berlin: de Gruyter.

Friedmann, Naama, Adriana Belletti, and Luigi Rizzi. 2021. Growing trees: The acquisition of the left periphery. *Glossa: a journal of general linguistics* 6:131.

Gomes, Antonio Almir Silva. 2013. Sanapaná, uma língua maskoy: Aspectos gramaticais. Doctoral dissertation, Universidade Estadual de Campinas, Campinas.

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

Hanson, Rebecca. 2000. Pronoun acquisition and the morphological feature geometry. In *Calgary Working Papers in Linguistics 22*, ed. by Susan Atkey, Jana Carson, and Michael Dobrovolsky, 1–14. Calgary: Department of Linguistics, University of Calgary.

Harbour, Daniel. 2014a. Paucity, abundance, and the theory of number. *Language* 90:185–229.

Harbour, Daniel. 2014b. Poor pronoun systems and what they teach us. *Nordlyd: Special Issue on Features* 41:125–143.

Harbour, Daniel. 2016. *Impossible Persons*. Cambridge, MA: The MIT Press.

Harley, Heidi, and Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78:482–526.

Noyer, Robert Rolf. 1992. Features, positions and affixes in autonomous morphological structure. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Rizzi, Luigi. 1993/4. Some remarks on linguistic theory and language development. *Language Acquisition* 3:371–393.

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

Smith, Norval. 2011. *Free Personal Pronoun System database*. Data Archiving and Networked Services (DANS).

Terenghi, Silvia. 2023. Missing person: Structure and change in Romance demonstratives. Doctoral dissertation, Utrecht University, Utrecht.

Tvica, Seid. 2017. Agreement and verb movement: The Rich Agreement Hypothesis from a typological perspective. Doctoral dissertation, Universiteit van Amsterdam, Amsterdam.

Vanden Wyngaerd, Guido. 2018. The feature structure of pronouns: A probe into multidimensional paradigms. In *Exploring Nanosyntax*, ed. by Lena Baunaz, Liliane Haegeman, Karen De Clercq, and Eric Lander, 277–304. Oxford/New York: Oxford University Press.

Silvia Terenghi
s.terenghi@uu.nl