Disentangling morphomic splits in Limbu

Berthold Crysmann — LLF, CNRS & U Paris Cité
Baptiste Loreau Unger — ENS Paris-Saclay & U Paris Cité

In this talk, we shall examine patterns of syncretism in the system of participant marking in Limbu (van Driem, 1987), a Kiranti language spoken in eastern Nepal by 180,000 people. Similar to other Kiranti languages, such as Athpare (Ebert, 1997b) or Camling (Ebert, 1997a), Limbu verbs inflect for their core arguments, corresponding to S(ole), A(gent) and P(atient) roles. Participants are marked for number (singular, dual, plural) and person (1,2,3), including an inclusive/exclusive distinction for first person non-singular. Marking is predominantly suffixal, with only a few prefixal markers for number, person, and negation.

On the one hand, participant marking in Limbu appears relatively transparent: in case of combination, participants tend to be marked individually, although cases of portmanteau marking do exist. Adding to the transparency, person and number distinctions for each participant are often marked separately by discrete markers. Moreover, the system of participant marking is largely the same across different tenses (non-past vs. past) or polarity.

On the other hand, this transparency contrasts with a number of syncretism patterns that affect different parts of the paradigm in different ways (cf. Table 1 for reference). Besides almost complete neutralisation of second person number contrasts in the $2>1$ and $1>2$ cells[1] (see §1.3), we also find partial neutralisation of the dual/plural distinction for third person A and third person P (see §1.2). This syncretism differs for A and P roles, providing an instance of divergent bidirectional syncretism in the terminology of Stump (2001). The third type of syncretism that complicates the system can be observed with allomorphic variation of person/number markers in different tenses (or polarities), giving rise to what we shall call "pseudo-Paninian" splits (see §1.1).

Table 1: Limbu person marking[2] (based on conjugation lists in van Driem, 1987, 368-374)

↓ A \ P →	1SG	1DE	1PE	1DI	1PI	2SG	2DU	2PL	3SG	3DU	3PL
1SG						-nɛ	-nɛ-tchi-ŋ	-n-i-ŋ	-u-ŋ | -ʔɛ / -paŋ	-u-ŋ-si-ŋ | -ʔɛ-n-chi-n / -paŋ-si-ŋ	→
1DE						←	-nɛ-tchi-ge	→	-s-u-ge	-s-u-si-ge	→
1PE						↙	↓	↘	-u-m-be / -mʔna	-u-m-si-m-be / -mʔna-si	→
1DI									a- -s-u	a- -s-u-si	→
1PI									a- -u-m	a- -u-m-si-m	→
2SG	kɛ- -ʔɛ / kɛ- -aŋ	↑	↗						kɛ- -u	kɛ- -u-si	→
2DU	←	a-gɛ-	→						kɛ- -s-u	kɛ- -s-u-si	→
2PL	↙	↓	↘						kɛ- -u-m	kɛ- -u-m-si-m	→
3SG	-ʔɛ / -aŋ	-si-ge	-i-ge	a- -si	a-	kɛ-	kɛ- -si	kɛ- -i	-u	-u-si	→
3DU	↑	↑	↑	↑	↑	↑	↑	↑	-s-u	-s-u-si	→
3PL	mɛ- -ʔɛ / mɛ- -aŋ	mɛ- -si-ge	mɛ- -i-ge	a-m- -si	a-m-	kɛ-m-	kɛ-m- -si	kɛ-m- -i	mɛ- -u	mɛ- -u-si	→
S →	-ʔɛ / -aŋ | -paŋ, -aŋ	-si-ge	-i-ge / -mʔna	a- -si	a-	kɛ-	kɛ- -si	kɛ- -i	-0	-si	mɛ-

[1] We use the (standard) $m > n$ notation to denote any participant with person/number features m acting on a patient with features n.

[2] Cells with allomorphy conditioned by tense or polarity are divided up into 2 by 2 subtables, with non-past at the top, past at the bottom, affirmative on the left and negative on the right. Regular tense and polarity marking has been omitted from the paradigm in the interest of readability (largely reproducing the non-past affirmative paradigm).

We use arrows to represent syncretism between adjacent cells, without necessarily implying any directionality.

1 Syncretism in Limbu participant marking

1.1 "Pseudo-Paninian" splits

The first case of syncretism patterns we shall discuss is witnessed in at least two places in the paradigm in Table 1: one involving the 2>1 paradigm (centre left in Table 1), the other involving exponents of first plural exclusive across different tenses and roles (cf. Table 2a).

Let us start with the 2>1 case. When taken in isolation, it just looks like your standard Paninian split: the *a-gɛ-* prefix serves to express all cells of this 3 by 3 sub-paradigm, effectively neutralising number distinctions, while there is a special circumfixal form for the 2SG>1SG cell that functions as an override. This form itself is peculiar: the *kɛ-* prefix also serves as a second-person marker notably in the 2>3 transitive and the 2 intransitive subparadigms. Similarly, the suffix, which features the two allomorphic variants *-ʔɛ* (NPST) and *-aŋ* (PST), can also be found in the 3>1SG cells and the 1SG cell of the intransitive paradigm. Thus, we are faced with the paradoxical situation that what functions locally as an override in the 2>1 subparadigm actually corresponds to more general forms used elsewhere in the expression of second and first person participants.

Table 2: Schematic representation of pseudo-Paninian splits

(a) First plural exclusive

Role	A>3SG		S		3SG>P	
Tense	NPST	PST	PST	NPST	PST	NPST
1PE	-u-m-be	**-mʔna**	**-mʔna**	-i-ge	-i-ge	-i-ge
2PL	kɛ- -u-m	kɛ- -u-m	kɛ- -i	kɛ- -i	kɛ- -i	kɛ- -i
1PI	a- -u-m	a- -u-m	a- -ɛ	a-	a- -ɛ	a-

(b) Dual/plural (*-si* vs. *mɛ-*)

Role	A>1/2	A>3	S	P
3DU	mɛ-	-si	-si	-si
3PL	mɛ-	mɛ-	mɛ-	-si

A highly similar behaviour can be observed for marking of first plural exclusive: looking at the 1PE column in both the transitive and intransitive paradigms, it appears that *-mʔna* is a specific portmanteau override in the past intransitive paradigm for the otherwise regular *-i-ge*, the latter being composed of the exclusive marker *-ge/-be* and the plural marker *-i*. Both exclusive and plural markers are attested in other areas of the full paradigm as well, such as the first exclusive dual (1DE) and plural (1PE) rows (*-ge/-be*) and the 2PL column (*-i*).

However, if we look more closely at the distribution of *-mʔna*, we find it in the 1PE>3 cells as well, where it is the past tense allomorph of *-u-m-be*. Again, each of these three markers is fairly general, marking third person P (*-u*), first/second plural A (*-m*), and first person exclusive (*-ge/-be*). Given the syncretism of *-mʔna* across S and A roles, it turns out to be more general than what we expect of a simple Paninian override: while it is indeed more specific than its competitors in most respects, combining past with first person exclusive plural, its role specification, viz. A or S (= "nominative"), is actually not more specific than either *-i* (S or P = "absolutive") or *-m* (A). The nature of the split is shown schematically in Table 2a: as one can easily discern, the split is neither fully natural (or balanced), nor fully Paninian. However, merely considering it as morphomic misses the clean separation of the (green) *-u-m* and the (red) *-i* cells. Furthermore, if Paninian competition can be invoked, it will be possible to maintain a natural description of the competitors *-u-m-be* and *-i-ge* in terms of the general properties expressed by their constituent formatives.

We shall argue that the issue with pseudo-Paninian splits can be resolved in this case by generalising the shared properties of *-mʔna* into a more abstract common rule type, yet expand this rule type into two rules that are individuated for the specific argument role (S vs. A).

1.2 Neutralisation of dual/plural

Although number marking in Limbu generally distinguishes dual and plural, there are regions in the paradigm in Table 1 where this distinction is effectively neutralised: most obviously,

for third person P participants, *-si* functions as a mere non-singular marker. In the 3>1/2 cells, the dual/plural distinction is equally neutralised for A participants, now featuring *mɛ-* as the exponent of non-singular third person A. However, in situations where the contrast between dual and plural is maintained, as e.g. for third person intransitive S, *mɛ-* serves to mark plural, whereas *-si* expresses dual. One way to picture this situation is in terms of divergent bidirectional syncretism (Stump, 2001) where the dual marker takes on expression of plural in the third person P cells, and the plural marker is extended to expression of dual cells in the third person A cells.

As depicted in Table 2b, the syncretism of *-si* and *mɛ-* in third person gives rise to a pattern of two interlocking L shapes. Thus, when taken in isolation, this looks like a "balanced" morphomic split where neither of the two markers can receive a straightforward natural characterisation, yet none of the two can be considered a default or an override either. The picture changes, however, once we include the full range of exponents for non-singular number: as it turns out, there is no other marker that uniquely encodes dual, but there are other markers (*-m,-i*) that specifically encode plural. As a consequence, it is safe to regard *-s(i)/-tchi* as a non-singular marker that only gets restricted to dual by virtue of (Paninian) competition with a dedicated plural marker. This is in line with the fairly wide distribution of *-si*: according to van Driem (1987), the alternation between *-si*, *-s* and *-tchi* is in most cases a mere phonologically conditioned allomorphy. Under this perspective, *mɛ-* is an A/S third person non-singular marker with two specialised instances: ambiguity-preserving in 3>1/2 cells, and plural otherwise. In sum, we can resolve the split similar to the case of *-m?na* discussed in §1.1 above.

1.3 Neutralisation of number

A particular neutralisation pattern affects the cells with only speech act participants (2>1, 1>2). Person marking for 2>1 uses a combination of the role-independent markers for first (*a-*) and second person (*kɛ-*) participants. Person marking for 1>2, by contrast, is expressed by the portmanteau marker *-nɛ*, encoding a first person A acting on second person, which preempts the role-independent markers via Panini's principle. Number however remains entirely unmarked in the 2>1 cells and for P in the 1NSG>2 cells. These cells are not only clearly exceptional in Limbu, but also in other Kiranti languages such as Athpare or Camling, giving rise to cross-linguistic variation.

2 Towards a formal analysis

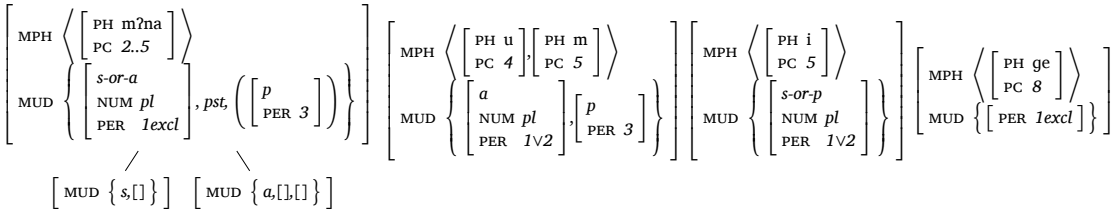
Previous formal analyses of Limbu participant marking have so far largely focused on the phenomenon of affix copying found with the *-ŋ* and *-m* markers (Zimmermann, 2012; Stump, 2022). Stump does provide a grammar fragment for part of the paradigm, but the intransitive and third person agent sub-paradigms are not covered. Thus, the specific issues of syncretism we are confronted with here have so far not been addressed.

The analysis we propose is couched in terms of Information-based Morphology (=IbM; Crysmann & Bonami, 2016), an inferential-realisational theory of inflection that implements a templatic view of morphotactics within a formalism based on inheritance hierarchies of rules.

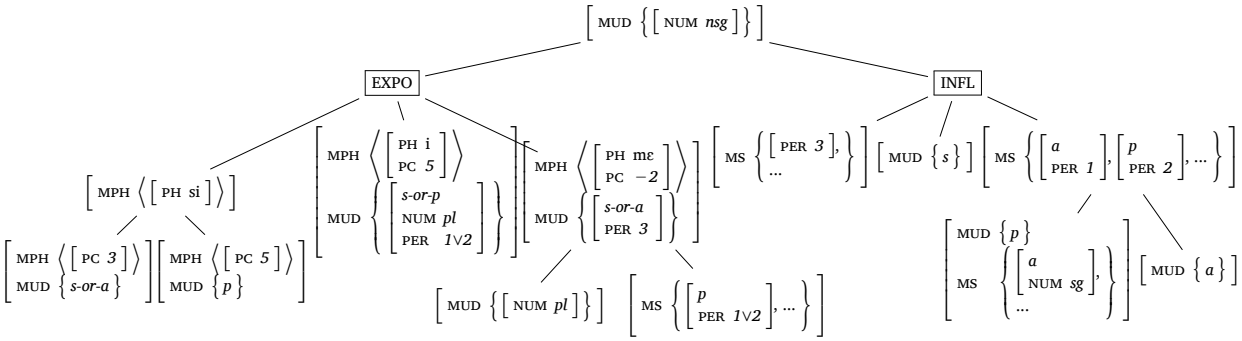
To start with, let us consider how the pseudo-Paninian split for *-m?na* can be captured: Figure 1a provides the relevant rules for portmanteau *-m?na* and its competitors *-u-m-be* and *-i-ge*. Crucially, the rules for *-m?na* are organised in a type hierarchy where the supertype generalises across the S and A cells. As can be easily verified, this supertype is neither more general nor more specific than its competitors or their combinations, since it is more informative with respect to tense, but less (*s-or-a* vs. *a*) or incommensurate (*s-or-a* vs. *s-or-p*) with respect to role. By providing subtypes for *-m?na*, however, we can individuate the general constraints to the specific roles (*a* vs. *s*), such that *-m?na* can serve as a true Paninian override in these contexts.

Turning to non-singular marking (cf. Figure 1b for a partial set of rules), there are two issues that need to be solved: first, constrain the macro-distribution of inflectional marking across the

paradigm (cf. §1.3), and second, orchestrate the competition between exponents (cf. §1.2). To address the former, we provide partial rules in the INFL dimension that constrain non-singular marking to two regular areas (expression of S/presence of a third person participant), as well as exceptional marking in the $1>2$ area. Rules of exponence in the EXPO dimension are cross-classified with these constraints on inflectedness, accounting for the absence of non-singular marking in the $2>1$ cells and the restricted distribution of *-si* and *-i* in the $1>2$ cells. With respect to exponence, we regard *-si* as the most general marker of non-singular, since it can be found in all three persons and all three roles (S, A, P). In case *-si* does not surface, there is either a (plural) competitor, or else no expression of number altogether. The rules for *-si* do not by themselves disambiguate between dual and plural, but they do positionally distinguish role (cf. Swahili; Crysmann & Bonami, 2016). Disambiguation of number arises by competition with markers that are either inherently plural (like *-i*, *-m* or plural inclusive ∅), or are specialised to plural in the relevant cells (*mɛ-*). Presence of non-singular *mɛ-* in the $3>1/2$ cells is equally derived by Panini's principle, given that the only non-singular competitor (*-si*) is more general, not bearing any person specification.



(a) *-m?na* and its competitors



(b) Dual/plural marking (*-si* vs. *mɛ-* and *-i*)

Figure 1: Rule hierarchies

To conclude, pseudo-Paninian splits and exceptional neutralisation of number marking in Limbu highlight the usefulness of underspecification and cross-classification in hierarchies of inflectional rules, to better encode and reconcile conflicting generalisations in complex morphological systems.

References

Crysmann, Berthold & Olivier Bonami. 2016. Variable morphotactics in Information-based Morphology. *Journal of Linguistics* 52(2). 311–374.

van Driem, George. 1987. *Limbu language*. Berlin: Mouton de Guyter.

Ebert, Karen. 1997a. *Camling (Chamling)*. München/Newcastle: Lincom Europa.

Ebert, Karen. 1997b. *A grammar of Athpare*. München/Newcastle: Lincom Europa.

Stump, Gregory T. 2001. *Inflectional morphology. A theory of paradigm structure*. Cambridge: Cambridge University Press.

Stump, Gregory T. 2022. *Morphotactics: A rule-combining approach*. Cambridge University Press.

Zimmermann, Eva. 2012. Affix copying in Kiranti. In Enrico Boone, Kathrin Linke & Maartje Schulpen (eds.), *Proceedings of ConSOLE XIX*, 343–367. Leiden: Leiden University.