



**HAL**  
open science

# The Opacity-Compactness Tradeoff: Morphomic Features for an Economical Account of Khaling Verbal Inflection

Géraldine Walther, Guillaume Jacques, Benoît Sagot

► **To cite this version:**

Géraldine Walther, Guillaume Jacques, Benoît Sagot. The Opacity-Compactness Tradeoff: Morphomic Features for an Economical Account of Khaling Verbal Inflection. 16th International Morphology Meeting (IMM 16), May 2014, Budapest, Hungary. hal-01114854

**HAL Id: hal-01114854**

<https://inria.hal.science/hal-01114854v1>

Submitted on 3 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## The Opacity-Compactness Tradeoff:

### Morphomic Features for an Economical Account of Khaling Verbal Inflection

GÉRALDINE WALTHER<sup>1,2</sup> GUILLAUME JACQUES<sup>3</sup> BENOÎT SAGOT<sup>4</sup>

1. DDL (CNRS, U. Lyon Lumière) 2. CRLAO (CNRS, Inalco)

3. LLF (CNRS, U. Paris Diderot) 4. ALPAGE (Inria, U. Paris Diderot)

geraldine.walther@univ-lyon2.fr rgyalrongskad@gmail.com benoit.sagot@inria.fr

Morphomic features (Aronoff, 1994) have been argued to allow for more elegant and more economical descriptions of inflectional systems (Bonami and Boyé, 2010). However, morphomic accounts are opaque from the morphosyntactic point of view and thus require an additional mapping between the morphomic features and the actual morphosyntactic property sets (MPS) expressed by the inflected forms (Stump, 2006) — something which is not needed if the MPSs are realised directly. Yet, no measures have been proposed so far to quantitatively assess the tradeoff between the lower complexity achieved by relying on opaque morphomic features and the additional need for an explicit feature mapping.

In this paper, we show that the descriptive economy achieved by capturing the inherent morphomic structure of inflectional systems can effectively be assessed through dedicated compactness measures.

We present two competing analyses of Khaling verbal inflection based on original data from Jacques *et al.* (2012). These two analyses will be called descriptions A and B hereafter. Khaling is a highly inflectional language of the Kiranti subbranch of the Sino-Tibetan language family, spoken by about 15,000 people in Eastern Nepal. Verbs realise up to 572 verb forms, among which 300 realis forms. Due to numerous syncretisms, these 300 realis forms reduce to only 100 distinct forms.

Among our two descriptions, DESCRIPTION A uses realisation rules based on standard MPS (such as A1S.P2S). DESCRIPTION B comprises rules realising morphomic features (such as B, C, D, etc. in Table 2) and their corresponding feature mapping functions (such as B: 1DU.INCL). We show that the paradigm structure of Khaling verbs is best described using morphomic features of the type proposed for Nepali by Bonami and Boyé (2010) together with another morphomic feature related to the direct/inverse marking system based on the language’s empathy hierarchy, as commonly used for describing Algonquian languages (Silverstein, 1976; Zúñiga, 2006).

We show that if one ignores the direct/inverse marking (i.e., the presence of the inverse prefix  $\text{ʔi-}$ ), the resulting additional syncretisms decrease the number of distinct forms from above mentioned 100 to only 68. These additional syncretisms are illustrated by the coloring in Table 1 for the non-past positive subparadigm. Additional morphomic feature mappings, such as posing a unique feature F for all forms expressing a P=2SG, then allows for further syncretic reduction. Table 2 shows how for the non-past positive subparadigm alone, the number of 75 distinct MPSs can be collapsed into 22 morphomic feature bundles such as single feature F or feature bundle  $X>1.2SG$ . Among these 22 morphomic feature bundles, 12 are complex features, but 10 are simple features, thus greatly simplifying the the expression of form realisation rules within the grammar (*lô:p-nu* becomes simply the realisation of feature K instead of that of features  $3PL>3$ ,  $3DU>3DU$ ,  $3DU>3PL$ ,  $2SG>3PL$  plus the inverse marked  $2SG>3PL$ ).

| ↓A   P→  | 1SG           | 1DU.INCL | 1DU.EXCL | 1PL.INCL   | 1PL.EXCL   | 2.SG    | 2DU      | 2PL        | 3SG       | 3DU        | 3PL        |
|----------|---------------|----------|----------|------------|------------|---------|----------|------------|-----------|------------|------------|
| 1SG      |               |          |          |            |            | loðm-ne | loðm-su  | loðm-nu    | lob-u     | lob-u-su   | lob-u-nu   |
| 1DU.INCL |               |          |          |            |            |         |          |            |           | løp-i      |            |
| 1DU.EXCL |               |          |          |            |            | ʔi-loop | ʔi-løp-i | ʔi-loðm-ni |           | løp-u      |            |
| 1PL.INCL |               |          |          |            |            |         |          |            |           | loop-ki    |            |
| 1PL.EXCL |               |          |          |            |            | ʔi-loop | ʔi-løp-i | ʔi-loðm-ni |           | loop-ka    |            |
| 2SG      | ʔi-loðm-ŋA    |          |          |            |            |         |          |            | ʔi-lô:b-u | ʔi-lô:p-su | ʔi-lô:p-nu |
| 2DU      | ʔi-loðm-ŋA-su |          |          |            |            |         |          |            |           | ʔi-løp-i   |            |
| 2PL      | ʔi-loðm-ŋA-nu |          |          |            |            |         |          |            |           | ʔi-loðm-ni |            |
| 3SG      | ʔi-loðm-ŋA    |          |          |            |            |         |          |            | lô:b-u    |            |            |
| 3DU      | ʔi-loðm-ŋA-su | ʔi-løp-i | ʔi-løp-u | ʔi-loop-ki | ʔi-loop-ka | ʔi-loop | ʔi-løp-i | ʔi-loðm-ni |           | lø:p-su    |            |
| 3PL      | ʔi-loðm-ŋA-nu |          |          |            |            |         |          |            |           |            | lø:p-nu    |

**Table 1.** Positive non-past paradigm for the Khaling verb LOP ‘catch’, represented with standard features

Our two descriptions have been implemented on 167 verbs, producing 50,100 verb forms. We have measured the descriptions in terms of descriptive economy following Sagot and Walther (2011). This measure relies on the information theoretical concept of Minimum Description Length (MDL) (Rissanen, 1984). The measure allows in particular for showing the amount of information within different parts of the de-

| X   | X>I        | I>X      | B     | C     | D      | E      | F   | G     | H       | I     | J      | K      |
|-----|------------|----------|-------|-------|--------|--------|-----|-------|---------|-------|--------|--------|
| 2SG | loðm-ŋA    | loðm-nɛ  |       |       |        |        |     |       |         |       |        |        |
| 2DU | loðm-ŋA-su | loðm-su  |       |       |        |        |     |       |         |       |        |        |
| 2PL | loðm-ŋA-nu | loðm-nu  | løp-i | løp-u | løp-ki | løp-ka | løp | løp-i | loðm-ni | løb-u | løp-su | løp-nu |
| 3S  | loðm-ŋA    | lob-u    |       |       |        |        |     |       |         |       |        |        |
| 3DU | loðm-ŋA-su | lob-u-su |       |       |        |        |     |       |         |       |        |        |
| 3PL | loðm-ŋA-nu | lob-u-nu |       |       |        |        |     |       |         |       |        |        |

**Table 2.** Positive non-past paradigm for the Khaling verb LOP ‘catch’, reorganised after having introduced morphomic feature mappings E.g., B: 1DU.INCL, F: P=2SG, I: 2SG>3SG or 3SG>3SG, ...

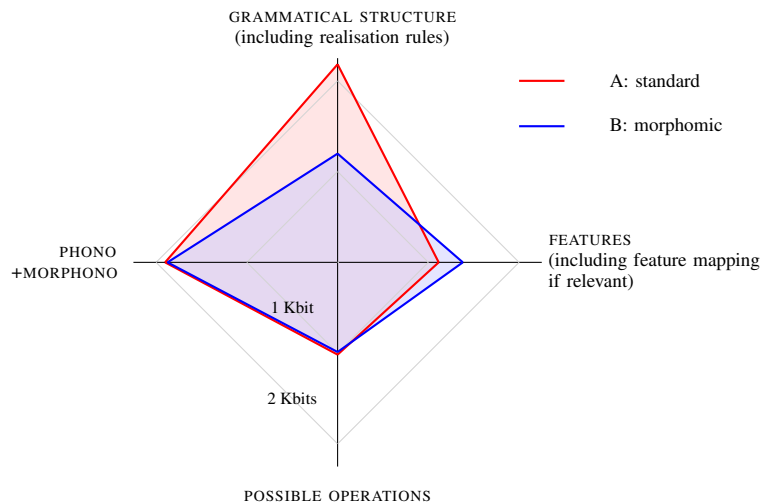
scription, such as regular morphonological rules at morph boundaries (“phono+morphono”), the types of phonological “operations” applied by the realisation rules, the information comprised within the “feature” specification and the feature mapping rules, and the general “structure” of the description, including the complexity of the realisation rules or the specified inflection classes.

In our experiment, the measure shows that the length of standard description A (with realisation rules realising transparent fully specified MPSs) is 6.2 Kbits, while the length of morphomic description B (realising above mentioned morphomic features in the rule sets coupled with separate inverse marking and morphomic mapping functions) is 5.4 Kbits.

These results show that using morphomic features makes for a description that is by 12% more compact than the non-morphomic description. In particular, the additional cost of feature mapping functions (here formalised with a mechanism similar to rules of *Paradigm Linkage* in PFM (Stump, 2006)) does not lead to a higher overall description length (due to the simplification of the realisation rules themselves). As illustrated by Figure 1, the additional information needed to stipulate the mapping between the morphomic features and the actual realised morphosyntactic property sets appears to be a more than reasonable tradeoff compared to the reduced length of the general description’s structure.

Combined with supplementary morphomic features, direct/inverse marking as traditionally used as a distinctive feature within descriptive linguistics thus proves to be quantitatively relevant for describing some morphological systems, as it can lead to formulating maximally compact descriptions.

**Acknowledgments** This work was conducted as part of the research operation LR4.11 which belongs to the Strand 6 “Language Resources” of the LabEx EFL (ANR-10-LABX-0083) funded by ANR.



**Fig. 1.** Comparative description lengths for two accounts of Khaling verbal inflection, breakdown on grammar components

## References

- ARONOFF, M.. 1994. *Morphology by Itself: Stems and Inflectional Classes*. Linguistic Inquiry Monographs. MIT Press.
- BONAMI, OLIVIER AND GILLES BOYÉ. 2010. Opaque paradigms, transparent forms in nepali conjugation. Workshop On Theoretical Morphology 5.
- JACQUES, GUILLAUME, AIMÉE LAHAUSSOIS, BOYD MICHAILOVSKY, AND DHAN BAHADUR RAI. 2012. An overview of Khaling verbal morphology. *Language and linguistics*, **13.6**, 1095–1170.
- RISSANEN, J.. 1984. Universal coding, information, prediction, and estimation. *IEEE Transactions on Information Theory*, **30(4)**, 629–636.
- SAGOT, BENOÎT AND GÉRALDINE WALTHER. 2011. Non-canonical inflection : data, formalisation and complexity measures. In C. MAHLOW AND M. PIOTROWSKI, Eds., *Systems and Frameworks in Computational Morphology*, volume 100, p. 23–45, Zurich, Switzerland: Springer.
- SILVERSTEIN, MICHAEL. 1976. Hierarchy of Features and Ergativity. In R. M. DIXON, Ed., *Grammatical Categories in Australian Languages*.
- STUMP, GREGORY T.. 2006. Heteroclis and paradigm linkage. *Language*, **82**, 279–322.
- ZÚÑIGA, FERNANDO. 2006. *Deixis and Alignment - Inverse systems in indigenous languages of the Americas*. Amsterdam: Benjamins.