



**HAL**  
open science

# The implicative structure of Asama verb paradigms

Dimitri Lévêque, Thomas Pellard

► **To cite this version:**

Dimitri Lévêque, Thomas Pellard. The implicative structure of Asama verb paradigms: A quantitative study of segmental and suprasegmental alternations. *Morphology*, 2023, 33 (3), pp.261–286. 10.1007/s11525-023-09410-x . hal-04151819

**HAL Id: hal-04151819**

**<https://hal.science/hal-04151819>**

Submitted on 9 Jul 2023

**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 implicative structure of Asama verb paradigms

## A quantitative study of segmental and suprasegmental alternations

Dimitri Lévêque\*

Thomas Pellard\* ✉ [thomas.pellard@cnrs.fr](mailto:thomas.pellard@cnrs.fr)

\*CNRS-EHESS-INALCO, CRLAO

---

**Abstract** Formal and computational linguistics can enhance descriptive linguistics of endangered languages by providing them with precise models and quantitative perspectives. We exemplify the benefits of such an approach with the case of Asama’s verb inflectional morphology. We show that a Word-and-Paradigm framework can provide interesting insights and allow for both the identification and the quantification of the sources of uncertainty in the implicative relations within Asama’s verb paradigms. We describe Asama’s verb morphology by considering whole forms rather than exponents only, and we factor its alternation patterns in two types: segmental alternations and suprasegmental alternations. Measures of Shannon’s conditional entropy are then used to estimate the respective contributions of these factors to the complexity of the system. Suprasegmental alternations turn out to be the major source of uncertainty in implicative relations, and vowel length and tone alternations cannot be treated separately but strongly interact. We also show how the principal parts of the inflectional system can be determined with conditional entropy measures of  $n$ -ary implicative relations.

**Keywords** Word-and-Paradigm, Implicative Morphology, entropy, paradigms, inflectional morphology, Asama, Ryukyuan, endangered languages

## 1 Introduction

Despite their many benefits, the methods and tools of computational linguistics are still seldom put to use for describing endangered, low-resource languages.<sup>1</sup> Here, we show how the description of the inflectional verb morphology of Asama, an endangered and underdescribed Ryukyuan language spoken in Southern Japan, can benefit from a formal and implemented model. Asama’s complex inflectional system can be accounted for in an insightful manner by adopting a Word-and-Paradigm framework, and a computational implementation allows for a quantitative perspective on its complexity and its sources.

The purpose of this paper is thus twofold: to provide a precise description of Asama’s inflectional verb morphology, and to investigate the sources of uncertainty that lie at the core of the inflectional system. First, we take advantage of a Word-and-Paradigm theoretical framework such as Implicative Morphology and provide a description of the inflectional system and a classification of verbal lexemes based on their patterns of alternation, highlighting the existence of morphomic zones within the paradigms. We show that this system can be described with two parallel sets of patterns, i.e. segmental alternations and suprasegmental alternations (tone and vowel length), each of which displays inflection classes and thus form together a cross-classifying system of different paradigmatic layers (Brown & Hippisley 2012, Baerman, Brown & Corbett 2017, Beniamine 2018, Parker & Sims 2020). We then

---

<sup>1</sup>See however Jacques et al. (2012), Snoek et al. (2014), Crysmann (2016), Harrigan et al. (2017), Pellard & Yamada (2017), and Chen, Park & Schwartz (2020) or Wilmoth & Mansfield (2021) for interesting exceptions.

compute complexity measures of the different sources of uncertainty in morphological alternations and investigate the relative informativity of implicative relations and stem phonology as well as their interaction.<sup>2</sup> We argue that generalisations need to be made on the basis of full word-forms instead of only alternating parts since non-alternating parts also contain important information for predicting other inflected forms.

## 2 Data sources and descriptive preliminaries

Asama is an endangered variety of Amami, a Northern Ryukyuan language of the Japonic family (Pellard 2015), spoken in the Asama village (Figure 1) on Tokunoshima Island, Kagoshima prefecture, Japan. It is spoken only by a few hundred speakers, at different levels of fluency, and it is not transmitted anymore to younger generations (there are no speakers under forty).

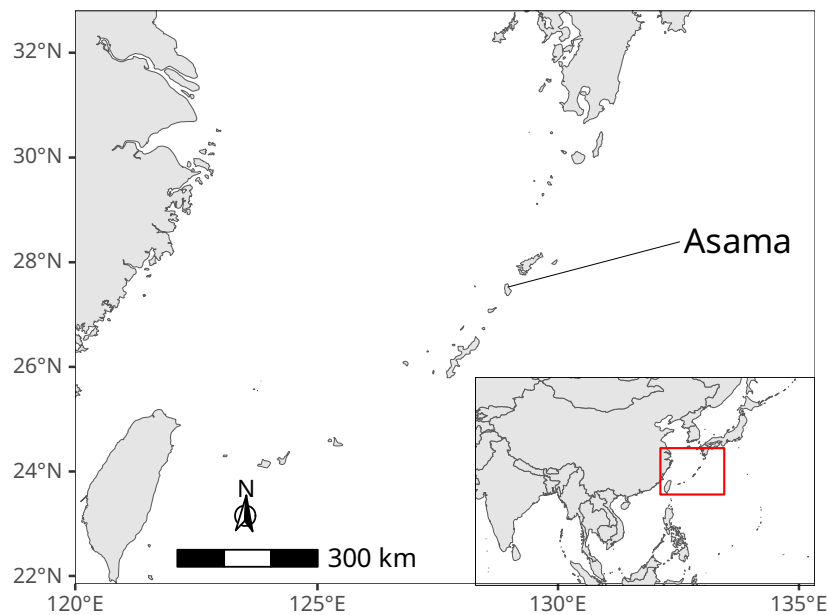


Figure 1: Location of Asama

This study is based on a corpus of 999 verb paradigms.<sup>3</sup> All the data used in this study are from a single speaker, Takahiro Okamura, born in 1936, and were either collected in the field by the first author between 2018 and 2020, or come from published sources (Uwano 2001, Okamura et al. 2009).

Asama differs from most other Japonic languages, even other Amami varieties,<sup>4</sup> by its peculiar system of suprasegmental features that interact with verb inflection. In Asama, all phonological words must include at least one heavy, i.e. bimoraic, syllable, which is defined as a syllable containing a long vowel, a diphthong, or a consonantal coda (either a nasal underspecified for place of articulation or the first part of a geminate consonant). This constraint leads to alternations between short and long vowels in inflected forms.

<sup>2</sup>See also the study on Russian noun inflection by Guzmán Naranjo (2020) and on Pitjantjatjara and Yankunytjatjara by Wilmoth & Mansfield (2021) that also consider different conditioning factors of inflection pattern.

<sup>3</sup>Verbs with a non-past form consisting of a single syllable, such as  $\text{NJUI}^{\text{LH}}$  'to boil' or  $\text{j'UI}^{\text{H}}$  'to say', are very few in number and are highly irregular, and they have thus been excluded from the present study. Two other lexemes,  $\text{KAMJUI}^{\text{LH}}$  'to eat' and  $\text{SIBUJUI}^{\text{H}}$  'to squeeze', have also been excluded due to contradictory data.

<sup>4</sup>See Niinaga (2015, 2014), van der Lubbe & Tokunaga (2015), and Yokoyama (Tokunaga) (2017).

Asama is also characterized by a word-tone system where melodies are mapped onto whole word forms (Uwano 1977, Hayata 1995, Lawrence 2015, Uwano 2016), and there is a strong interdependence between tone pattern and syllable weight, as will be seen in the following sections. Inflected forms of verbs show four distinctive tone patterns: H, LH, HL and LHL. Hereafter, tone patterns are indicated as abstract melodies written as exponents at the end of the word form rather than with diacritics denoting their surface realisations.

- (1) a. H: all syllables are high, except the last one if it is light;  
e.g. *íkjúi* ‘to go’ (NPST)
- b. LH: if the word is shorter than five morae, only the last mora is high, otherwise the syllable including the fifth mora is high, as well as all the following syllables, except for the last one if it is light;  
e.g. *òmòjúi* ‘to think’ (NPST)
- c. HL: the first syllable is high, and all the following ones are low;  
e.g. *áadùrà̀n* ‘to knit’ (PROG.NEG)
- d. LHL: only one given syllable in the middle of the word is high (depending on the length of the form and the inflectional class of the lexeme);  
e.g. *tùutùrà̀n* ‘to take’ (PROG.NEG)

Asama is typologically close to other Japonic languages, but its verb inflectional morphology departs from the simple and almost canonically agglutinative type exhibited by Japanese. Like some other Ryukyuan languages, such as Yonaguni (Pellard & Yamada 2017), Asama is characterized by a comparatively more complex and opaque morphology, exhibiting several non-canonical phenomena (in the sense of Corbett 2009) such as segmental and suprasegmental (tone and vowel length) stem allomorphy, systematic overabundance and multi-functional exponents.

This study is based on the partial paradigm of eleven cells listed in Table 1. The full paradigm of Asama’s verb morphology contains 38 cells (without counting suffix combinations), but their realised forms can all be simply derived from those of the eleven cells and add no uncertainty to the system (Lévêque 2021).

Table 1: Partial verb paradigm

Form	Abbreviation
Non-past	NPST
Converb 2	CVB2
Negation	NEG
Direct imperative	DIMP
Indirect imperative	IIMP
Converb	CVB
Desiderative	DES
Past	PST
Sequential	SEQ
Progressive	PROG
Progressive negative	PROG.NEG

### 3 Non-canonical phenomena as sources of uncertainty

#### 3.1 Overabundance

Asama's verb inflection exhibits several non-canonical phenomena, such as several cases of systematic overabundance, where two different realisations are possible for the same paradigm cell. This is the case for the progressive past, for which two variants exist that differ in the location of the long vowel: either in the stem or in the suffix (2).

- (2) a. 'to begin' {*hazimiit-ut-an*<sup>H</sup>, *hazimit-uut-an*<sup>H</sup>}  
b. 'to be born' {*maariit-ut-an*<sup>H</sup>, *maarit-uut-an*<sup>H</sup>}

#### 3.2 Suffix allomorphy

Second, the same morphosyntactic feature is sometimes realised by different exponents, depending on the presence of other morphosyntactic features. For instance, the negative is marked by *-an* in non-past forms, but with *-adaat-* when followed by the past suffix *-i* (3).

- (3) a. 'to buy' NPST *koor-an*<sup>H</sup> : PST *koor-adaat-i*<sup>H</sup>  
b. 'to read' NPST *jum-an*<sup>LH</sup> : PST *jum-adaat-i*<sup>LH</sup>  
c. 'to think' NPST *omor-an*<sup>LH</sup> : PST *omor-adaat-i*<sup>LH</sup>

#### 3.3 Multi-functional exponents

Third, Asama also exhibits multi-functional exponents, which express radically different morphosyntactic properties despite being formally identical. This is the case of the suffix *-i*, which can mark either the indirect imperative or the sequential (4), or the suffix *-an*, which can mark either the negative or the past (5), depending on which stem they attach to.

- (4)
- |           |                             |                             |
|-----------|-----------------------------|-----------------------------|
|           | IIMP                        | SEQ                         |
| 'to read' | <i>juum-i</i> <sup>LH</sup> | <i>juud-i</i> <sup>LH</sup> |
| 'to meet' | <i>oor-i</i> <sup>H</sup>   | <i>oot-i</i> <sup>H</sup>   |
- (5)
- |           |                             |                             |
|-----------|-----------------------------|-----------------------------|
|           | NEG                         | PST                         |
| 'to read' | <i>jum-an</i> <sup>LH</sup> | <i>jud-an</i> <sup>LH</sup> |
| 'to meet' | <i>oor-an</i> <sup>H</sup>  | <i>oot-an</i> <sup>H</sup>  |

#### 3.4 Stem alternations

Non-canonical phenomena affecting the suffixal part of inflected forms are nevertheless few in Asama, and the major source of morphological opacity lies in stem alternations, which affect all verbs, as well as all forms within a paradigm.

Table 2 gives partial paradigms for four representative verbs, with colors highlighting the distribution of stem alternations. Each verb exhibits at least four distinct stem shapes, but their exact number and their distribution differs between verbs, which greatly increases the amount of information needed to account for the whole set of cells.

Several segmental and suprasegmental factors contribute to the complexity of stem alternations. First, the final part of stems is subject to segmental alternations (Table 3, §4.2.1). Some verbs also exhibit vowel length alternations, where a vowel appears as short in some cells but as long in others (Table 4).

Table 2: Partial paradigm for four verbs

	‘to sell’	‘to knit’	‘to talk’	‘to play’
NPST	<i>uj-<b>ui</b></i> <sup>H</sup>	<i>amj-<b>ui</b></i> <sup>LH</sup>	<i>hanasj-<b>ui</b></i> <sup>LH</sup>	<i>asibj-<b>ui</b></i> <sup>H</sup>
CVB2	<i>ur-ug99si</i> <sup>H</sup>	<i>amj-ug99si</i> <sup>LH</sup>	<i>hanasj-ug99si</i> <sup>LH</sup>	<i>asibj-ug99si</i> <sup>H</sup>
NEG	<i>ur-an</i> <sup>H</sup>	<i>am-an</i> <sup>LH</sup>	<i>hanasj-an</i> <sup>LH</sup>	<i>asib-an</i> <sup>H</sup>
DIMP	<i>ur-oo</i> <sup>H</sup>	<i>am-oo</i> <sup>LH</sup>	<i>hanasj-oo</i> <sup>LH</sup>	<i>asib-oo</i> <sup>H</sup>
IIMP	<i>ur-ii</i> <sup>H</sup>	<i>aam-i</i> <sup>HL</sup>	<i>hanaasj-i</i> <sup>LH</sup>	<i>asiib-i</i> <sup>H</sup>
CVB	<i>u-i</i> <sup>H</sup>	<i>am-ii</i> <sup>LH</sup>	<i>hanas-ii</i> <sup>LH</sup>	<i>asib-ii</i> <sup>H</sup>
DES	<i>u-icjaahai</i> <sup>H</sup>	<i>am-icjaahai</i> <sup>LH</sup>	<i>hanas-icjaahai</i> <sup>LH</sup>	<i>asib-icjaahai</i> <sup>H</sup>
PST	<i>ut-an</i> <sup>H</sup>	<i>ad-an</i> <sup>LH</sup>	<i>hanacj-an</i> <sup>LH</sup>	<i>asid-an</i> <sup>H</sup>
SEQ	<i>ut-ii</i> <sup>H</sup>	<i>aad-i</i> <sup>HL</sup>	<i>hanaacj-i</i> <sup>LH</sup>	<i>asiid-i</i> <sup>H</sup>
PROG	<i>ut-<b>ui</b></i> <sup>LH</sup>	<i>aad-<b>ui</b></i> <sup>HL</sup>	<i>hanaacj-<b>ui</b></i> <sup>LHL</sup>	<i>asid-<b>ui</b></i> <sup>LH</sup>
PROG.NEG	<i>ut-uuran</i> <sup>LH</sup>	<i>aad-uran</i> <sup>HL</sup>	<i>hanaacj-uran</i> <sup>LHL</sup>	<i>asid-uuran</i> <sup>LH</sup>

Table 3: Segmental alternations (alternating parts are indicated in bold)

	‘to sell’	‘to knit’	‘to talk’	‘to play’
NPST	<i>uj-<b>ui</b></i> <sup>H</sup>	<i>amj-<b>ui</b></i> <sup>LH</sup>	<i>hanasj-<b>ui</b></i> <sup>LH</sup>	<i>asibj-<b>ui</b></i> <sup>H</sup>
CVB2	<i>ur-ug99si</i> <sup>H</sup>	<i>amj-ug99si</i> <sup>LH</sup>	<i>hanasj-ug99si</i> <sup>LH</sup>	<i>asibj-ug99si</i> <sup>H</sup>
NEG	<i>ur-an</i> <sup>H</sup>	<i>am-an</i> <sup>LH</sup>	<i>hanasj-an</i> <sup>LH</sup>	<i>asib-an</i> <sup>H</sup>
DIMP	<i>ur-oo</i> <sup>H</sup>	<i>am-oo</i> <sup>LH</sup>	<i>hanasj-oo</i> <sup>LH</sup>	<i>asib-oo</i> <sup>HL</sup>
IIMP	<i>ur-ii</i> <sup>H</sup>	<i>aam-<b>i</b></i> <sup>H</sup>	<i>hanaasj-<b>i</b></i> <sup>LH</sup>	<i>asiib-<b>i</b></i> <sup>H</sup>
CVB	<i>u-i</i> <sup>H</sup>	<i>am-ii</i> <sup>LH</sup>	<i>hanas-ii</i> <sup>LH</sup>	<i>asib-ii</i> <sup>H</sup>
DES	<i>u-icjaahai</i> <sup>H</sup>	<i>am-icjaahai</i> <sup>LH</sup>	<i>hanas-icjaahai</i> <sup>LH</sup>	<i>asib-icjaahai</i> <sup>H</sup>
PST	<i>ut-an</i> <sup>H</sup>	<i>ad-an</i> <sup>LH</sup>	<i>hanacj-an</i> <sup>LH</sup>	<i>asid-an</i> <sup>H</sup>
SEQ	<i>ut-ii</i> <sup>H</sup>	<i>aad-<b>i</b></i> <sup>HL</sup>	<i>hanaacj-<b>i</b></i> <sup>LH</sup>	<i>asiid-<b>i</b></i> <sup>H</sup>
PROG	<i>ut-<b>ui</b></i> <sup>LH</sup>	<i>aad-<b>ui</b></i> <sup>HL</sup>	<i>hanaacj-<b>ui</b></i> <sup>LHL</sup>	<i>asid-<b>ui</b></i> <sup>LH</sup>
PROG.NEG	<i>ut-uuran</i> <sup>LH</sup>	<i>aad-uran</i> <sup>HL</sup>	<i>hanaacj-uran</i> <sup>LHL</sup>	<i>asid-uuran</i> <sup>LH</sup>

Table 4: Vowel length alternations (alternating parts are indicated in bold)

	‘to sell’	‘to knit’	‘to talk’	‘to play’
NPST	<i>uj-<b>ui</b></i> <sup>H</sup>	<i>amj-<b>ui</b></i> <sup>LH</sup>	<i>hanasj-<b>ui</b></i> <sup>LH</sup>	<i>asibj-<b>ui</b></i> <sup>H</sup>
CVB2	<i>ur-ug99si</i> <sup>H</sup>	<i>amj-ug99si</i> <sup>LH</sup>	<i>hanasj-ug99si</i> <sup>H</sup>	<i>asibj-ug99si</i> <sup>H</sup>
NEG	<i>ur-an</i> <sup>H</sup>	<i>am-an</i> <sup>LH</sup>	<i>hanasj-an</i> <sup>LH</sup>	<i>asib-an</i> <sup>H</sup>
DIMP	<i>ur-oo</i> <sup>H</sup>	<i>am-oo</i> <sup>LH</sup>	<i>hanasj-oo</i> <sup>LH</sup>	<i>asib-oo</i> <sup>H</sup>
IIMP	<i>ur-ii</i> <sup>H</sup>	<i>aam-<b>i</b></i> <sup>HL</sup>	<i>hanaasj-<b>i</b></i> <sup>LH</sup>	<i>asiib-<b>i</b></i> <sup>H</sup>
CVB	<i>u-i</i> <sup>H</sup>	<i>am-ii</i> <sup>LH</sup>	<i>hanas-ii</i> <sup>LH</sup>	<i>asib-ii</i> <sup>H</sup>
DES	<i>u-icjaahai</i> <sup>H</sup>	<i>am-icjaahai</i> <sup>LH</sup>	<i>hanas-icjaahai</i> <sup>LH</sup>	<i>asib-icjaahai</i> <sup>H</sup>
PST	<i>ut-an</i> <sup>H</sup>	<i>ad-an</i> <sup>LH</sup>	<i>hanacj-an</i> <sup>LH</sup>	<i>asid-an</i> <sup>H</sup>
SEQ	<i>ut-ii</i> <sup>H</sup>	<i>aad-<b>i</b></i> <sup>HL</sup>	<i>hanaacj-<b>i</b></i> <sup>LH</sup>	<i>asiid-<b>i</b></i> <sup>H</sup>
PROG	<i>ut-<b>ui</b></i> <sup>LH</sup>	<i>aad-<b>ui</b></i> <sup>HL</sup>	<i>hanaacj-<b>ui</b></i> <sup>LHL</sup>	<i>asid-<b>ui</b></i> <sup>LH</sup>
PROG.NEG	<i>ut-uuran</i> <sup>LH</sup>	<i>aad-uran</i> <sup>HL</sup>	<i>hanaacj-uran</i> <sup>LHL</sup>	<i>asid-uuran</i> <sup>LH</sup>

Table 5: Tone alternations (in bold)

	'to sell'		'to knit'		'to talk'		'to play'	
NPST	<i>uj-ui</i>	<b>H</b>	<i>amj-ui</i>	<b>LH</b>	<i>hanasj-ui</i>	<b>LH</b>	<i>asibj-ui</i>	<b>H</b>
CVB2	<i>ur-ug<sup>99</sup>si</i>	<b>H</b>	<i>amj-ug<sup>99</sup>si</i>	<b>LH</b>	<i>hanasj-ug<sup>99</sup>si</i>	<b>LH</b>	<i>asibj-ug<sup>99</sup>si</i>	<b>H</b>
NEG	<i>ur-an</i>	<b>H</b>	<i>am-an</i>	<b>LH</b>	<i>hanasj-an</i>	<b>LH</b>	<i>asib-an</i>	<b>H</b>
DIMP	<i>ur-oo</i>	<b>H</b>	<i>am-oo</i>	<b>LH</b>	<i>hanasj-oo</i>	<b>LH</b>	<i>asib-oo</i>	<b>H</b>
IIMP	<i>ur-ii</i>	<b>H</b>	<i>aam-i</i>	<b>HL</b>	<i>hanaasj-i</i>	<b>LH</b>	<i>asiib-i</i>	<b>H</b>
CVB	<i>u-i</i>	<b>H</b>	<i>am-ii</i>	<b>LH</b>	<i>hanas-ii</i>	<b>LH</b>	<i>asib-ii</i>	<b>H</b>
DES	<i>u-icjaaha<sub>1</sub></i>	<b>H</b>	<i>am-icjaaha<sub>1</sub></i>	<b>LH</b>	<i>hanas-icjaaha<sub>1</sub></i>	<b>LH</b>	<i>asib-icjaaha</i>	<b>H</b>
PST	<i>ut-an</i>	<b>H</b>	<i>ad-an</i>	<b>LH</b>	<i>hanacj-an</i>	<b>LH</b>	<i>asid-an</i>	<b>H</b>
SEQ	<i>ut-ii</i>	<b>H</b>	<i>aad-i</i>	<b>HL</b>	<i>hanaacj-i</i>	<b>LH</b>	<i>asiid-i</i>	<b>H</b>
PROG	<i>ut-ui</i>	<b>LH</b>	<i>aad-ui</i>	<b>HL</b>	<i>hanaacj-ui</i>	<b>LHL</b>	<i>asid-ui</i>	<b>LH</b>
PROG.NEG	<i>ut-uuran</i>	<b>LH</b>	<i>aad-uran</i>	<b>HL</b>	<i>hanaacj-uran</i>	<b>LHL</b>	<i>asid-uuran</i>	<b>LH</b>

Another factor is the tone alternations seen between different paradigm cells (Table 5). However, the distribution of long vowels and that of tones are not independent, and the two features need to be analysed together (§4.2.2).

Tables 2 to 5 show that a clearer view of Asama’s verb morphology obtains once the stem alternations are factorised into three different types: segmental alternations, vowel length alternations, and tonal alternations. It is then possible to investigate the complexity of Asama’s verb morphology in more detail by examining to what extent the different stems are predictable from each other, and how each alternation type contributes to the overall complexity of the system.

## 4 Morphological analysis and implicative structure

### 4.1 Theoretical framework

Most descriptive works on the morphology of Ryukyuan languages (Niinaga 2014, Shimoji 2017, Yokoyama (Tokunaga) 2017) are morpheme-based and consider inflectional forms to be built through the linear concatenation of morphemes, i.e. minimal meaningful units of form, with some morphophonological readjustments. However, the very concept of morpheme poses well-known theoretical and practical problems (Robins 1959, Matthews 1972, Hockett 1987, Blevins 2016, Bauer 2016), and in the case of Asama, a morpheme-based approach is problematic when dealing with suprasegmental stem alternations, which constitute the major locus of morphological complexity. Under a morpheme-based approach, it would be necessary to introduce many highly idiosyncratic morphophonological rules in order to get surface forms from underlying ones, not only for segmental alternations but especially for suprasegmental alternations. It is doubtful that such an approach would lead to any gain in terms of generalisations, adequacy, simplicity or economy. Instead of focusing on syntagmatic structure and relations only, we thus adopt a Word-and-Paradigm approach.<sup>5</sup>

We thus focus on paradigmatic alternations, rather than on the syntagmatic combination of isolated elements. We examine in particular the implicative relations that hold between inflected forms and the relative difficulty of predicting the realisation of a paradigm cell from that of another cell or set of cells. Our approach builds on the existence of an implicative structure within inflectional paradigms

<sup>5</sup>See Blevins (2016) for a recent general overview of the topic.

(Wurzel 1989), and is thus directly related to the *Paradigm Cell Filling Problem* (Ackerman, Blevins & Malouf 2009, Ackerman & Malouf 2013), i.e. “[w]hat licenses reliable inferences about the inflected (and derived) surface forms of a lexical item?” (Ackerman, Blevins & Malouf 2009: p. 54), as well as the problem of finding an inflectional system’s principal parts (Finkel & Stump 2007, 2009, Stump & Finkel 2013). We show that many of the alternations described in §3.4 can be efficiently described in terms of implicative relations, and that describing inflectional paradigms can often be reduced to the description of their implicative structure.

## 4.2 Morphological analysis

The verb morphology of Asama forms a bipartite system (Beniamine 2018: pp. 106–122) that can be described with two different sets of alternations, each associated with a distinct set of inflectional classes. The first set concerns segmental alternations, and the second one suprasegmental alternations, i.e. vowel length and tone. This factorisation is justified by the fact that the two sets of alternations are independent one of each other and should thus be treated separately.

### 4.2.1 Segmental alternations

The first set of alternations concerns segmental phenomena. For all verbs, the stem-final consonant of inflected forms is subject within a paradigm to segmental alternations that differ across inflectional classes, while the suffixal part remains stable in most cases, apart for some automatic morphophonological alterations we describe in §4.2.2. Disregarding suprasegmental alternations as in Table 3, it is thus possible to divide all inflectional forms into three parts: a non-alternating root part  $R$ , an alternating stem consonant  $\Sigma$  (potentially zero, indicated in bold in Table 3), and a suffixal part, e.g. for *amj-ui*<sup>LH</sup> ‘to knit-NPST’ in Table 3,  $R = a$ ,  $\Sigma = \mathbf{mj}$ , and the suffixal part is *-ui*.

On the basis of their alternation patterns, the different cells of the paradigm can be partitioned into the five isomorphic zones  $T_1$ ,  $T_2$ ,  $T_3$ ,  $T_4$ , and  $T_5$  defined in (6), i.e. sets of cells that share the same realisation of the alternating consonant. This partition of the paradigm is morphomic, i.e. it is purely internal to morphology, and does not depend on semantic, syntactic or phonological criteria (Aronoff 1994), even if the division in classes is partially correlated with the identity of the consonant. For each zone  $T_i$ , there exists a realisation  $\Sigma_i$  of the alternating consonant depending on which class a lexeme belongs to (Table 6).

- (6)  $T_1$ : non-past, interrogative  
 $T_2$ : converb 2  
 $T_3$ : desiderative, converb  
 $T_4$ : negative, hortative, imperative  
 $T_5$ : progressive, past, sequential

For instance, consider the sequential form  $w^{\prime}iizji^H$  of the verb ‘to swim’. Its alternating consonant is  $zj$ , and the sequential form belongs to the zone  $T_5$ . Since there are three different inflectional classes for which  $\Sigma_5 = zj$ , if we try to infer the corresponding non-past form, which belongs to  $T_1$ , we observe that for  $T_1$ ,  $\Sigma_1 \in \{nj, gj, kj\}$ . We can thus conclude that for the non-past, the segmental form of the verb ‘to swim’ is either  $w^{\prime}ijnj-ui$ ,  $w^{\prime}igj-ui$ , or  $w^{\prime}ikj-ui$ .



Table 6: Segmental alternations

Class	$\Sigma_1$	$\Sigma_2$	$\Sigma_3$	$\Sigma_4$	$\Sigma_5$
b	bj	bj	b	b	d
m	mj	mj	m	m	d
n	nj	nj	n	nj	zj
g	gj	gj	g	g	zj
k1	kj	kj	k	k	cj
k2	kj	kj	k	k	ccj
k3	kj	kj	k	k	zj
kk	kkj	kkj	kk	kk	ccj
s1	sj	sj	s	sj	cj
s2	sj	sj	s	sj	sj
ss	ssj	ssj	ss	ssj	ccj
c	cj	cj	c	t	ccj
cc	ccj	ccj	cc	tt	ccj
aou1	j	r	$\emptyset$	r	t
aou2	j	r	$\emptyset$	r	tt
aou3	j	r	$\emptyset$	r	cj
i1	j	r	r	r	t
i2	j	r	r	r	cj
i3	j	r	r	r	ccj
si1	j	r	r	r	t
si2	j	r	r	r	cj
si3	j	r	r	r	ccj
Ci1	j	r	$\emptyset$	r	t
Ci2	j	r	$\emptyset$	r	tt
Ci3	j	r	$\emptyset$	r	ccj

## 4.2.2 Suprasegmental alternations

Though segmental alternations could be described with a classical morpheme-based approach, which posits that rules apply to concatenated strings of underlying morphemes (notwithstanding several problems), segmenting an inflectional form into parts poses problems when dealing with suprasegmental alternations. Tone patterns apply to a whole word form, not to a single part of the form, such as the stem or the suffix, and vowel length alternations can affect either the stem or the suffix. Such problems can be avoided by abstracting alternation patterns from whole inflected forms rather than by trying to segment forms and isolate stems and exponents (see also Beniamine, Bonami & Luís 2021).<sup>6</sup>

On the basis of suprasegmental alternations, verb paradigms can be partitioned into four areas of full interpredictability, where knowledge of one cell of the area provides full knowledge of all the other cells of the area (i.e. isomorphic zones). Like for segmental alternations, such a partition of the paradigm is morphomic, as can be inferred from the non-exhaustive list of forms belonging to those four areas provided in (7).

- (7) A: non-past, negative, hortative  
 B: converb, desiderative, prohibitive  
 C: sequential, imperative  
 D: progressive, preparative

This partition is represented in the columns of Table 7. In Table 8 a single cell has been chosen to represent each area as an illustration in order to avoid accumulating abstract letters. The choice of the representative cell is arbitrary, and any other cell belonging to the same area could have been chosen. This is what Stump & Finkel (2013: p. 42) call a *distillation*, i.e. a single set of morphosyntactic properties among several isomorphic property sets.

Tables 7 and 8 represent the different alternation patterns across the different zones and their distillation, and the different inflectional classes they define. For instance, class III can be defined as the class exhibiting a LH tone for the non-past (zone A) and the converb (zone B), and a HL tone as well as a long vowel in the root for the sequential (zone C) and the negative progressive (zone D).

For example, the sequential form *tudii*<sup>H</sup> of ‘to fly’ corresponds the pattern  $R\Sigma_5 i:$ <sup>H</sup> of class II1 in Table 8 since we can identify its root part  $R = tu$  without vowel lengthening, its alternating consonant  $\Sigma_5 = d$ , its suffix  $i:$ , and its tone H.

Only three rules need to be further applied to those patterns in order to obtain the surface forms (8). They are mostly the result of general phonological well-formedness constraints and thus have no impact on the predictability of the morphological system.

- (8) a. Super-heavy syllables that comprise more than two morae are not allowed, and there is no vowel lengthening if  $R$  already ends with a heavy syllable, or if  $R$  ends with a consonant in coda position;  
 if  $R$  ends in  $CVV$ ,  $R: \rightarrow CVV$  ( $*CVVV$ ), and if  $R$  ends in  $CVC$ ,  $R: \rightarrow CVC$  ( $*CVVC$ )
- b. The vowel  $i$  at the beginning of a suffix can surface as  $i$  due to a ban on the occurrence of  $i$  after the coronal consonants  $s$ ,  $z$ ,  $c$ ,  $t$ ,  $d$ , and  $r$ ;  
 $i \rightarrow i / \{s, z, c, t, d, r\} \_$
- c. The vowel  $i$  at the beginning of a suffix is dropped after a central vowel  $i$  or  $\text{ɔ}$ , which undergoes compensatory lengthening instead;  $ii$  and  $\text{ɔ}i$  sequences are rare if attested at all in the native lexicon of Asama.  
 $i \rightarrow i / i, \text{ɔ} \_$ ,  $i \rightarrow \text{ɔ} / \text{ɔ} \_$

<sup>6</sup>Here we base our analyses on phonological forms akin to lexical representations, rather than on surface phonetic forms or underlying morphophonological ones.

Table 7: Patterns of suprasegmental alternations; A, B, C, and D represent the four zones of full inter-predictability; I1, I2, etc. are suprasegmental inflectional classes;  $\Sigma$  stands for the alternating consonant described in §4.2.1;  $X$  and  $Y$  represent affixes or part of affixes (necessary because of the possibility of a vowel lengthening inside the affixal part); when two different forms appear in a cell, this indicates they are in free variation

Class	A	B	C	D
I1	$R\Sigma X^H$	$R\Sigma X^H$	$R:\Sigma X^H$	$R\Sigma X:Y^H$
I2	$R\Sigma X^H$	$R\Sigma X^H$	$R:\Sigma X^H$	$R:\Sigma XY^H, R\Sigma X:Y^H$
II1	$R\Sigma X^H$	$R\Sigma X^H$	$R\Sigma X:^H$	$R\Sigma X:Y^{LHL}$
II2	$R\Sigma X^H$	$R\Sigma X^H$	$R:\Sigma X^H$	$R\Sigma X:Y^{LHL}$
III	$R\Sigma X^{LH}$	$R\Sigma X^{LH}$	$R:\Sigma X^{HL}$	$R:\Sigma X^{HL}$
IV	$R\Sigma X^{LH}$	$R\Sigma X^{LH}$	$R:\Sigma X^{LH}$	$R:\Sigma X^{LHL}$
V	$R:\Sigma X^{HL}$	$R:\Sigma X^{HL}$	$R:\Sigma X^{HL}$	$R:\Sigma X^{HL}$

Table 8: Patterns of suprasegmental alternations (distillation)

Class	A: {NPST}	B: {CVB}	C: {SEQ}	D: {PROG.NEG}
I1	$R\Sigma_1 ui^H$	$R\Sigma_3 i:^H$	$R:\Sigma_5 \dot{i}^H$	$R\Sigma_5 uuran^H$
I2	$R\Sigma_1 ui^H$	$R\Sigma_3 i:^H$	$R:\Sigma_5 \dot{i}^H$	$R\Sigma_5 uuran^H, R:\Sigma_5 uran^H$
II1	$R\Sigma_1 ui^H$	$R\Sigma_3 i:^H$	$R\Sigma_5 \dot{i}^H$	$R\Sigma_5 uuran^{LHL}$
II2	$R\Sigma_1 ui^H$	$R\Sigma_3 i:^H$	$R:\Sigma_5 \dot{i}^H$	$R\Sigma_5 uuran^{LHL}$
III	$R\Sigma_1 ui^{LH}$	$R\Sigma_3 i:^{LH}$	$R:\Sigma_5 \dot{i}^{HL}$	$R:\Sigma_5 uran^{HL}$
IV	$R\Sigma_1 ui^{LH}$	$R\Sigma_3 i:^{LH}$	$R:\Sigma_5 \dot{i}^{LH}$	$R:\Sigma_5 uran^{LHL}$
V	$R:\Sigma_1 ui^{HL}$	$R:\Sigma_3 \dot{i}^{HL}$	$R:\Sigma_5 \dot{i}^{HL}$	$R:\Sigma_5 uran^{HL}$

### 4.3 Whole forms in implicative structure

In this section, we provide justifications for the claim that it is necessary to take whole forms into account in order to describe Asama’s verb morphology in a consistent and insightful manner. We show that, in contradistinction with the approach taken by [Ackerman, Blevins & Malouf \(2009\)](#), it is necessary to consider whole realised forms and not only alternating parts or exponents. Otherwise several important generalisations can be missed since non-alternating parts of inflected forms also contain important information for predicting other forms.

#### 4.3.1 Segmental alternations

From a segmental point of view, the only alternating part in Asama’s inflected verb forms is the consonant  $\Sigma$  (§4.2.1). However, important generalisations can be drawn from the phonotactic structure of the non-alternating part  $R$  too, and some segmental classes of Table 6 differ in the phonotactic properties of their non-alternating part.

For example, lexemes with a non-alternating part  $R$  ending with a vowel have a non-past form  $Rjui$ , and their converb form can be either  $Ri$ ,  $R$ , or  $Rri$ . However, knowing the identity of the last segment

of  $R$  reduces the number of possible forms, though not completely (9).<sup>7</sup>

- (9) a. If  $R$  ends in a back vowel, then  $[\text{NPST} : R\text{jui} \Rightarrow \text{CVB} : R\text{i}]$   
 e.g. ‘to laugh’:  $R = \text{waro}$ ,  $[\text{NPST} : \text{warojui}^{\text{H}} \Rightarrow \text{CVB} : \text{waroi}^{\text{H}}]$
- b. If  $R$  ends in  $\text{ɔ}$ , then  $[\text{NPST} : R\text{jui} \Rightarrow \text{CVB} : R:]$   
 e.g. ‘to move’:  $R = \text{mugɔ}$ ,  $[\text{NPST} : \text{mugɔjui}^{\text{H}} \Rightarrow \text{CVB} : \text{mugɔ}^{\text{H}}]$
- c. If  $R$  ends in  $\text{i}$ , then  $[\text{NPST} : R\text{jui} \Rightarrow \text{CVB} : \left\{ \begin{array}{l} R: \\ R\text{rii} \end{array} \right\}]$   
 e.g. ‘to rub’:  $R = \text{si}$ ,  $[\text{NPST} : \text{sijui}^{\text{LH}} \Rightarrow \text{CVB} : \text{siri}^{\text{LH}}]$   
 e.g. ‘to fall’:  $R = \text{uti}$ ,  $[\text{NPST} : \text{utijui}^{\text{H}} \Rightarrow \text{CVB} : \text{utii}^{\text{LH}}]$

The two first conditions in (9) follow from the rules stated in (8). The uncertainty of (9c) can be resolved by looking not only at the last vowel of  $R$ , which is uniformly  $\text{i}$ , but also at the preceding consonant (10).

- (10) if  $R$  ends in  $\text{i}$ ,
- a. and if the preceding consonant is either  $\text{c}$ ,  $\text{z}$  or  $\text{s}$ , then  
 $[\text{NPST} : R\text{jui} \Rightarrow \text{CVB} : R\text{rii}]$   
 e.g. ‘to rub’:  $R = \text{si}$ ,  $[\text{NPST} : \text{sijui}^{\text{LH}} \Rightarrow \text{CVB} : \text{siri}^{\text{LH}}]$
- b. otherwise,  $[\text{NPST} : R\text{jui} \Rightarrow \text{CVB} : R:]$   
 e.g. ‘to fall’:  $R = \text{uti}$ ,  $[\text{NPST} : \text{utijui}^{\text{H}} \Rightarrow \text{CVB} : \text{utii}^{\text{LH}}]$

This demonstrates the informative character of non-alternating parts for implications. Non-alternating parts can help resolve uncertainties that remain when considering exponents and alternating parts only.

### 4.3.2 Suprasegmental alternations

Little if any information about the suprasegmental class of a lexeme can be drawn from just looking at the alternating part in a paradigm cell. Considering whole forms is crucial since the prosodic weight and syllable structure of non-alternating parts can help resolving uncertainties (Table 9).

Table 9: Syllable structure and prosodic weight as information sources

Class	Weight of $R$	Syllable structure of $R$	$R$	NPST	Gloss
I1	$2\mu$	(C)VV, (C)VCV	<i>waro</i>	<i>warojui</i> <sup>H</sup>	‘to laugh’
I2	$\geq 3\mu$	any	<i>m’aari</i>	<i>m’aarijui</i> <sup>H</sup>	‘to be born’
II1	$\leq 2\mu$	(C)V, (C)VC	<i>tu</i>	<i>tubjui</i> <sup>H</sup>	‘to fly’
II2	$2\mu$	(C)VCV, (C)VN	<i>asi</i>	<i>asibjui</i> <sup>H</sup>	‘to play’
III	$\leq 2\mu$	(C)V, (C)VN, (C)VC	<i>u</i>	<i>umjui</i> <sup>LH</sup>	‘to ripen’
IV	any	$\neq (C^{\text{v}})VN, (C^{\text{v}})VC$	<i>tuu</i>	<i>tuujui</i> <sup>LH</sup>	‘to go through’
V	$2\mu$	$CV_iV_i$	<i>m’oo</i>	<i>m’oojui</i> <sup>HL</sup>	‘to be’ (honorific)

<sup>7</sup>We use the symbol  $\Rightarrow$  to denote implicative relationships, i.e. the implication that for any lexeme  $L$ , if the set of morphosyntactic features  $\phi$  is realised by  $x$ , then the set of morphosyntactic features  $\psi$  is realised by  $y$  is noted  $[\phi : x \Rightarrow \psi : y]$  (Bonami 2014: p. 86). Braces  $\{\dots\}$  indicate sets of alternative forms.

For instance, from a non-past form such as *tubjui*<sup>H</sup> ‘to fly’, two different possible sequential forms can be inferred, depending on whether the lexeme belongs to class II1 or to either class I1, I2, or II2 (11).

$$(11) \left[ \text{NPST} : R\Sigma_1 ui^H \Rightarrow \text{SEQ} : \left\{ \begin{array}{l} R\Sigma_5 ii^H \\ R:\Sigma_5 i^H \end{array} \right\} \right]$$

e.g.  $\left[ \text{NPST} : \textit{tubjui}^H \Rightarrow \text{SEQ} : \left\{ \begin{array}{l} \textit{tudi}^H \\ *\textit{tuudi}^H \end{array} \right\} \right]$

However, knowing that the non-alternating part  $R$  of *tubjui*<sup>H</sup> is *tu* and that its weight  $M(R)$  is thus less than two morae, we can resolve this uncertainty: the sequential form is *tudi*<sup>H</sup>, and the lexeme *TUBJUI*<sup>H</sup> thus belongs to class II1 (12).

$$(12) \left[ \text{NPST} : R\Sigma_1 ui^H \text{ and } M(R) \leq 2\mu \Rightarrow \text{SEQ} : R\Sigma_5 ii^H \right]$$

e.g.  $\left[ \text{NPST} : \textit{tubjui}^H \Rightarrow \text{SEQ} : \textit{tudi}^H \right]$

In Table 7, all patterns are thus described only by syllable structure and weight, as detailed in Table 9 for each class, with the exception of class IV where the nature of the segment filling the onset of the first syllable of the form must be taken into account.

## 5 Measuring uncertainty

### 5.1 Method

The complexity of an inflectional system can be estimated as its uncertainty, as measured by Shannon entropy (Ackerman, Blevins & Malouf 2009, Ackerman & Malouf 2013). The entropy  $H(X)$  of a random discrete variable  $X$  that can take the values  $x_1, x_2, \dots, x_n$ , each with a probability  $P(x_1), P(x_2), \dots, P(x_n)$ , is the measure in bits of the uncertainty of the outcome of  $X$  (Shannon 1948a,b) as defined in (13).

$$(13) H(X) = - \sum_{x \in X} P(x) \log_2 P(x)$$

Within the context of inflectional morphology, the variable  $X$  is a paradigm cell and its values are its different possible realisations across inflectional classes, which usually vary in their frequency and thus their probability. Entropy varies according to both the total number of possible outcomes and the imbalance in their relative frequencies: the higher the entropy, the higher the uncertainty, and an entropy of zero indicates the absence of uncertainty. The maximum entropy of a variable is reached when its  $n$  different outcomes are all equiprobable, i.e.  $0 \leq H(X) \leq \log_2 n$ .

However, when a speaker makes an inference about the realisation of paradigm cell, it is always based on prior knowledge of at least one other form, and we thus need to consider conditional entropy  $H(X | Y)$ , a measure of the uncertainty of  $X$  given knowledge of another variable  $Y$  (14). This gives us a measure of (lack of) interpredictability, i.e. the relative difficulty of predicting the realisation of a paradigm cell  $X$  from that of another cell or set of cells  $Y$ . A conditional entropy  $H(X | Y)$  lower than the simple entropy  $H(X)$  indicates that knowing the realisation of  $Y$  helps predicting that of  $X$ , and if  $H(X | Y)$  equals zero, then the realisation of  $X$  is completely determined by that of  $Y$ . On the other hand,  $H(X | Y)$  can never exceed  $H(X)$ , and if  $H(X | Y)$  equals  $H(X)$ , knowing the realisation of  $Y$  does not help predicting that of  $X$ . The simple conditional entropy  $H(X | Y)$  given a single cell  $Y$  for unary, i.e. one-to-one implications, can also easily be extended to cases of sets of cells  $H(X | Y, Z)$  for  $n$ -ary implications.

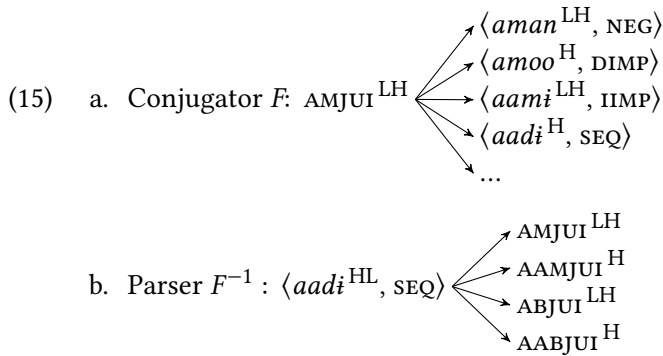
$$(14) \quad H(X | Y) = - \sum_{x \in X} \sum_{y \in Y} P(x, y) \log_2 P(x | y) = - \sum_{x \in X} \sum_{y \in Y} P(x, y) \log_2 \frac{P(x, y)}{P(x)}$$

Since we need to refer to whole inflected forms and thus to include both segmental and suprasegmental alternations, we first combined the information of Table 6 and Table 7. We then extracted a distillation of five cells which represent areas of systematic full interpredictability for both kinds of alternation. These five representative cells are listed in Table 10, which gives a description of Asama’s inflectional verb morphology in terms of alternation patterns (see also Blevins 2016, Bonami & Beniamine 2016).

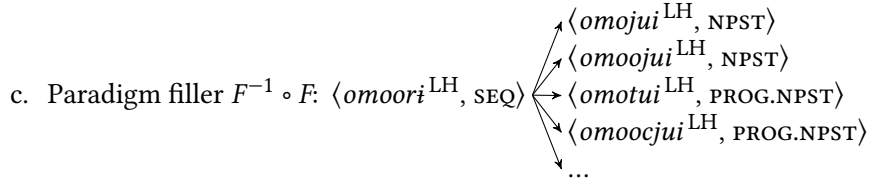
Table 10: Suprasegmental alternations and patterns of alternation; the highlighted pattern is discussed in text below

NPST	CVB	DIMP	SEQ	PROG.NEG
$R\Sigma_1 ui^H$	$R\Sigma_3 ii^H$	$R:\Sigma_4 i^H$	$R:\Sigma_5 i^H$	$R\Sigma_5 uuran^H$
$R\Sigma_1 ui^H$	$R\Sigma_3 ii^H$	$R:\Sigma_4 i^H$	$R:\Sigma_5 i^H$	$R:\Sigma_5 uran^H, R\Sigma_5 uuran^H$
$R\Sigma_1 ui^H$	$R\Sigma_3 ii^H$	$R\Sigma_4 ii^H$	$R\Sigma_5 ii^H$	$R\Sigma_5 uuran^{LHL}$
$R\Sigma_1 ui^H$	$R\Sigma_3 ii^H$	$R:\Sigma_4 i^H$	$R:\Sigma_5 i^H$	$R\Sigma_5 uuran^{LHL}$
$R\Sigma_1 ui^{LH}$	$R\Sigma_3 ii^{LH}$	$R:\Sigma_4 i^{HL}$	$R:\Sigma_5 i^{HL}$	$R:\Sigma_5 uran^{HL}$
$R\Sigma_1 ui^{LH}$	$R\Sigma_3 ii^{LH}$	$R:\Sigma_4 i^{LH}$	$R:\Sigma_5 i^{LH}$	$R:\Sigma_5 uran^{LHL}$
$R\Sigma_1 ui^{HL}$	$R\Sigma_3 ii^{HL}$	$R\Sigma_4 i^{HL}$	$R\Sigma_5 i^{HL}$	$R\Sigma_5 uran^{HL}$

In order to calculate conditional entropy, we implemented our morphological description as a set of finite-state transducers with the foma software (Hulden 2009).<sup>8</sup> We first implemented a conjugator  $F$  that for the abstract representation  $L$  of any lexeme gives its full paradigm, i.e. the set of all pairs  $\langle w, \sigma \rangle$  comprising a realised inflected form  $w$  and a set of morphosyntactic properties  $\sigma$  (Stump 2016). We inverted that transducer ( $F^{-1}$ ) to obtain a parser that relates any realised cell  $\langle w, \sigma \rangle$  to all possible abstract representations  $L$  of which  $\langle w, \sigma \rangle$  is a realised cell. We finally composed the two ( $F^{-1} \circ F$ ) to obtain a paradigm filler that from a realised cell  $\langle w, \sigma \rangle$  gives all the other cells of the paradigm, including all different possibilities in case of ambiguity (15).



<sup>8</sup>We thus follow the way set by an increasing number of works on well-described (Beniamine, Bonami & Luís 2021) or under-described (Jacques et al. 2012, Snoek et al. 2014, Crysmann 2016, Harrigan et al. 2017, Pellard & Yamada 2017, Chen, Park & Schwartz 2020) languages and acknowledge the “insufficiency of paper-and-pencil linguistics” (Karttunen 2006).



For all 999 lexemes in our data, we calculated the conditional entropy for all 25 two-cell combinations of the five representative cells. For any realised paradigm cell  $\langle w, \sigma \rangle$  of the distillation, our paradigm filler was used to generate all possible realisations of the four other cells without any other prior knowledge such as inflectional class membership. For instance, consider trying to infer the non-past form of ‘to think’ from its sequential form  $omooti^{LH}$ . Table 10 shows that  $omooti^{LH}$  instantiates the highlighted pattern  $R\Sigma_5 i^{LH}$ , and that it corresponds to a single pattern in the non-past column ( $R\Sigma_1 ui^{LH}$ ). Table 6 shows that the alternating consonant of the non-past form is  $j$  for all classes where the sequential form’s alternating consonant is  $t$ . The only uncertainty concerns vowel length since the long vowel  $oo$  in the sequential form could be either the result of lengthening or lexically stipulated and thus present in all forms of this lexeme. The non-past form could thus be either  $omojui^{LH}$  or  $omoojui^{LH}$ . Therefore, the conditional entropy  $H(NPST | SEQ)$  for ‘to think’ equals 1 bit (16).

$$(16) \quad H(NPST | SEQ) = -\left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2}\right) = 1$$

## 5.2 Frequency data

Entropy measures crucially depend on the relative type frequency of the different patterns, i.e. the variable  $P(x)$  in the definition given in (13). In most cases, disregarding frequency data thus leads to overestimating entropy. However, this is not always the case, since it also depends on the internal structure of the paradigms. For instance, consider the paradigm in Table 11, where there are two different forms for cell  $X$ ,  $a$  and  $b$ , and five different forms for cell  $Y$ , with four associated with  $a$  and only one associated with  $b$ . Conditional entropy in (17) is then higher ( $1.99 > 1.6$ ) when frequency is taken into account than when it is ignored (i.e. all frequencies equal 1).

Table 11: Entropy with and without frequency

Cell $X$	Cell $Y$	Frequency 1	Frequency 2
$a$	$c$	1	50
$a$	$d$	1	50
$a$	$e$	1	50
$a$	$f$	1	50
$b$	$g$	1	1

(17) a. Frequency 1:

$$\begin{aligned}
 H_1(Y | X) &= -\frac{1}{5} \log_2(1) - \frac{4}{5} \left( \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) \\
 &= \frac{8}{5} \\
 &= 1.6
 \end{aligned}$$

b. Frequency 2:

$$\begin{aligned}
 H_2(Y | X) &= -\frac{1}{201} \log_2(1) - \frac{200}{201} \left( 4 \frac{50}{200} \log_2 \frac{50}{200} \right) \\
 &= \frac{400}{201} \\
 &\approx 1.99 > H_1(Y | X)
 \end{aligned}$$

Table 12 and Table 13 give the number of occurrences and relative frequencies for respectively segmental and suprasegmental alternation patterns calculated on 999 Asama verb lexemes. The overall distribution is unbalanced, with a few alternation patterns covering most lexemes, and most patterns including only a few lexemes.

Table 12: Frequency of segmental alternation classes

Class	Occurrences	Frequency (%)	Class	Occurrences	Frequency (%)
b	16	1.60	aou1	245	24.52
m	72	7.21	aou2	6	0.60
n	1	0.10	aou3	4	0.40
g	35	3.50	i1	16	1.60
k1	99	9.91	i2	13	1.30
k2	1	0.10	i3	7	0.70
k3	1	0.10	si1	22	2.20
kk	6	0.60	si2	2	0.20
s1	114	11.41	si3	2	0.20
s2	4	0.40	Ci1	316	31.63
ss	1	0.10	Ci2	3	0.30
c	8	0.80	Ci3	3	0.30
cc	2	0.20	Total	999	100

Table 13: Frequency of suprasegmental alternation classes

Class	Occurrences	Frequency (%)
I <sub>1</sub>	228	22.82
I <sub>2</sub>	188	18.82
II <sub>1</sub>	75	7.51
II <sub>2</sub>	8	0.80
III	12	1.20
IV	485	48.55
V	3	0.30
Total	999	100



Table 14: Conditional entropy without frequency data; bright colors indicate a higher value than dark ones

$H(\text{Col}   \text{Row})$	NPST	CVB	IIMP	SEQ	PROG.NEG	Predictiveness
NPST		0	0	1.441	1.750	0.798
CVB	0.901		0.193	1.622	2.163	1.220
IIMP	0.975	0.534		1.439	2.100	1.262
SEQ	1.876	1.661	0.918		0.695	1.287
PROG.NEG	1.708	1.496	0.942	0		1.037
Predictability	1.365	0.923	0.513	1.125	1.677	1.121

Table 15: Conditional entropy with frequency data; bright colors indicate a higher value than dark ones

$H(\text{Col}   \text{Row})$	NPST	CVB	IIMP	SEQ	PROG.NEG	Predictiveness
NPST		0	0	0.200	0.228	0.107
CVB	0.949		0.063	0.266	0.598	0.469
IIMP	0.985	0.230		0.200	0.590	0.501
SEQ	1.381	1.075	0.410		0.423	0.822
PROG.NEG	1.247	0.953	0.410	0		0.653
Predictability	1.140	0.565	0.220	0.166	0.460	0.510

This discrepancy in the frequencies of the different patterns has a strong impact on the entropy of the system. Compare Table 14, which considers all types to be equiprobable, and Table 15, where the relative frequencies of both classes of alternations (segmental and suprasegmental) are taken into account. The predictability and predictiveness values are the mean values of the corresponding lines and columns. Including frequency data reduces the entropy values and thus increases predictability as expected, since the maximum entropy is reached when all types are equiprobable.

Importantly, not all cells are impacted in the same way, and there is no relation of proportionality between the values of Table 14 and those of Table 15. For example, the conditional entropy of the sequential given the non-past  $H(\text{SEQ} | \text{NPST})$  is more than 7 times greater without frequency data, while conditional entropy for the non-past given the progressive negative  $H(\text{NPST} | \text{PROG.NEG})$  is only 1.3 times greater. There are even two cases where entropy increases when frequencies are included:  $H(\text{NPST} | \text{CVB})$  and  $H(\text{NPST} | \text{IIMP})$ . In the following, all entropy values are calculated on the basis of the frequencies of Tables 12 and 13.

### 5.3 Sources of uncertainty

In §3, we showed that Asama morphology exhibited three different kinds of alternations: segmental alternations, vowel length alternations, and tone alternations. In the following, we examine those three factors separately in order to assess their respective contribution to the uncertainty of the realisations of paradigm cells. We focus here on unary implications, i.e. implications where there is only a single predictor cell (Bonami & Beniamine 2016).

### 5.3.1 Segmental alternations

Segmental alternations are one source of uncertainty in the realisation of Asama’s verb paradigms. Table 16 gives the proportion of entropy variation when removing suprasegmental alternations and retaining segmental alternations only.

Table 16: Proportion (%) of entropy variation in unary implications due to the removal of suprasegmental alternations; negative values indicate a decrease and positive values an increase in entropy (colors mapped onto a log-modulus transformed scale)

	NPST	CVB	IIMP	SEQ	PROG.NEG	Predictiveness
NPST		0	0	0	-12.3	-6.5
CVB	-100.0		-100.0	-24.8	-66.6	-78.7
IIMP	-100.0	-100.0		0	-66.1	-80.0
SEQ	-70.4	-62.0	-0.2		-100.0	-62.8
PROG.NEG	-67.2	-57.1	-0.2	0		-53.1
Predictability	-82.1	-63.9	-7.3	-9.6	-67.4	-64.3

We observe that for most cells, conditional entropy is globally much lower when only segmental alternations are retained (e.g.  $H(\text{CVB} \mid \text{SEQ}) = 0.409 \ll 1.075$ ), which means that segmental alternations do not play a major role in the uncertainty of verb paradigms as compared to suprasegmental alternations. Moreover, our initial distillation of five cells can be further reduced to two zones wherein conditional entropy is zero, i.e. wherein full interpredictability obtains: Zone 1, which includes the non-past, converb, and indirect imperative cells, and Zone 2, which includes the sequential and progressive negative cells (Table 18). The uncertainty between the two remaining zones can be said to be entirely due to suprasegmental alternations.

Looking back at (6) and Table 6, this means that the segmental realisations of  $\Sigma$  in cells belonging to our initial zones  $T_1$  to  $T_4$  are all interpredictable. Only zone  $T_5$  presents segmental realisations largely unpredictable from that of the other zones. These quantitative results are consistent with our initial preliminary observations.

When compared with the original conditional entropy values, those without suprasegmental alternations (Tables 16–17) provide another important piece of information: the mean predictiveness of the non-past form is not very different whether we take into account suprasegmental alternations (0.107) or not (0.1), i.e. disregarding suprasegmental features reduces predictiveness by only 6.5%. This indicates that once the non-past form is known, most of the remaining uncertainty concerns the segmental realisation, and that conversely the non-past has a high predictive power for suprasegmental alternations. It is however not the case of the converb and the imperative, and we can conclude that the non-past is a better predictor than those, no matter what is the source of uncertainty.

Examples in (18) illustrate the fact that when conditional entropy is reduced to zero when removing suprasegmental alternations, we can conclude that the uncertainty is due to suprasegmental features (here vowel length), and that segmental alternations (in bold) are fully predictable. On the other hand, examples in (19) illustrates the kind of unpredictable segmental alternations that increase the conditional entropy.

Table 17: Variations in conditional entropy values when removing sources of uncertainty

Predictor	Predicted	Original	Disregarding alternations of		
			Suprasegmental	Vowel length	Tones
NPST	CVB	0.000	0.000	0.000	0.000
	IIMP	0.000	0.000	0.000	0.022
	SEQ	0.200	0.200	0.198	0.224
	PROG.NEG	0.228	0.200	0.233	1.004
	mean	0.107	0.100	0.108	0.312
CVB	NPST	0.949	0.000	0.780	0.944
	IIMP	0.063	0.000	0.066	0.044
	SEQ	0.266	0.200	0.254	0.278
	PROG.NEG	0.598	0.200	0.288	0.987
	mean	0.469	0.100	0.172	0.573
IIMP	NPST	0.985	0.000	0.016	0.983
	CVB	0.230	0.000	0.002	0.216
	SEQ	0.200	0.200	0.198	0.201
	PROG.NEG	0.590	0.200	0.236	0.921
	mean	0.501	0.100	0.113	0.580
SEQ	NPST	1.381	0.409	0.428	1.386
	CVB	1.075	0.409	0.410	1.032
	IIMP	0.410	0.409	0.408	0.414
	PROG.NEG	0.423	0.000	0.040	0.735
	mean	0.822	0.306	0.322	0.892
PROG.NEG	NPST	1.247	0.409	0.428	1.291
	CVB	0.953	0.409	0.410	0.979
	IIMP	0.410	0.409	0.408	0.414
	SEQ	0.000	0.000	0.000	0.000
	mean	0.653	0.306	0.310	0.671
mean	NPST	1.140	0.204	0.237	1.151
	CVB	0.565	0.204	0.205	0.557
	IIMP	0.220	0.204	0.221	0.234
	SEQ	0.166	0.150	0.163	0.175
	PROG.NEG	0.460	0.150	0.199	0.912

Table 18: Impact of segmental alternations on the conditional entropy of unary implications (simplified)

$H(Col   Row)$	Zone 1	Zone 2	Predictiveness
Zone 1		0.200	0.200
Zone 2	0.409		0.409
Predictability	0.409	0.200	0.304

- (18) a.  $\left[ \text{IIMP} : m^{\text{aariiri}}^{\text{H}} \Rightarrow \text{NPST} : \left\{ \begin{array}{l} *m^{\text{aariijui}}^{\text{H}} \\ m^{\text{aarijui}}^{\text{H}} \end{array} \right\} \right]$  ‘to be born’
- b.  $\left[ \text{CVB} : \text{toi}^{\text{LH}} \Rightarrow \text{NPST} : \left\{ \begin{array}{l} \text{toojui}^{\text{LH}} \\ *tojui^{\text{LH}} \end{array} \right\} \right]$  ‘to pass’
- (19) a.  $\left[ \text{NPST} : k^{\text{imajui}}^{\text{H}} \Rightarrow \text{SEQ} : \left\{ \begin{array}{l} k^{\text{imaati}}^{\text{H}} \\ *k^{\text{imaacji}}^{\text{H}} \\ *k^{\text{imatti}}^{\text{H}} \end{array} \right\} \right]$  ‘to get decided’
- b.  $\left[ \text{PROG.NEG} : \text{izjii}^{\text{LH}} \Rightarrow \text{CVB} : \left\{ \begin{array}{l} \text{izjii}^{\text{LH}} \\ *izjii^{\text{LH}} \\ *izjikii^{\text{LH}} \end{array} \right\} \right]$  ‘to go out’

### 5.3.2 Vowel length alternations

Before studying the impact of suprasegmental features, it is important to note that vowel length and tone are not independent, as shown by the variation of entropy values when disregarding tones only (Table 19). While it is expected that removing a source of uncertainty would decrease the entropy values, we observe an increase in some cases instead. This means that the amount of information about vowel length decreases if tones are removed, and thus that vowel length and tones are not independent. The respective contributions of vowel length and tone to uncertainty cannot thus be accurately estimated independently.

Table 19: Proportion (%) of entropy variation in unary implications due to the removal of tone alternations; negative values indicate a decrease and positive values an increase in entropy (colors mapped onto a log-modulus transformed scale)

	NPST	CVB	IIMP	SEQ	PROG.NEG	Predictiveness
NPST		0	0	+12.0	+340.4	+191.6
CVB	-0.5		-30.2	+4.5	+65.1	+22.2
IIMP	-0.2	-6.1		+0.5	+56.1	+15.8
SEQ	+0.4	-4.0	+1.0		+73.8	+8.5
PROG.NEG	+3.5	+2.7	+1.0	0		+2.8
Predictability	+1.0	-1.4	+6.4	+5.4	+98.3	+18.7

Nevertheless, the increase in entropy observed when removing tones but not vowel length suggests that the contribution of tones to uncertainty is smaller than that of vowel length since the decrease in entropy due to removal of tones is out-weighted by additional uncertainties about vowel length. For instance,  $H(\text{PROG.NEG} \mid \text{NPST})$  equals 0.228 bits with tones but 1.004 bits without tones. This is confirmed by the decrease of entropy observed in Table 20, where vowel length but not tones is ignored. This shows that vowel length has an important contribution to the conditional entropy of unary implications.

Vowel length alternations are a cause of uncertainty because some inflectional forms undergo a process of automatic vowel lengthening, thus neutralising the distinction between short and long vowels. This makes it impossible to determine whether a long vowel in the non-alternating part of an inflected form is lexically specified or morphophonologically conditioned. In the first case, a long vowel appears throughout a lexeme’s paradigm, whereas in the second case it is restricted to a subset

Table 20: Proportion (%) of entropy variation in unary implications due to the removal of vowel length alternations; negative values indicate a decrease and positive values an increase in entropy (colors mapped onto a log-modulus transformed scale)

	NPST	CVB	IIMP	SEQ	PROG.NEG	Predictiveness
NPST		0	0	-1.0	+2.2	+0.9
CVB	-17.8		+4.8	-4.5	-51.8	-63.3
IIMP	-98.4	-99.1		-1.0	-60.0	-77.4
SEQ	-69.0	-61.9	-0.5		-90.5	-60.8
PROG.NEG	-65.7	-57.0	-0.5	0		-52.5
Predictability	-79.2	-63.7	+0.5	-1.8	-56.7	-59.8

of it. For example, all verbs have a long vowel in the non-alternating part of their indirect imperative form, but in the non-past form, the same vowel appears as long for some verbs and as short for some others (20). Thus, given the indirect imperative form *waroori*<sup>H</sup> ‘to laugh’, it is not possible to correctly infer that the long vowel is not lexically specified and that its non-past form is thus *warojui*<sup>H</sup> and not *\*waroojui*<sup>H</sup>. Conversely, it is not possible to know from the indirect imperative form *koori*<sup>H</sup> ‘to buy’ that the long vowel is lexically specified for that lexeme and that its non-past form is *koojui*<sup>H</sup> and not *\*kojui*<sup>H</sup>. This explains why the entropy value  $H(\text{IIMP} \mid \text{NPST})$  is greater than  $H(\text{NPST} \mid \text{IIMP})$  ( $0.985 > 0$ ) and why it greatly decreases (0.016, i.e. -98.4%) when vowel length is removed (Tables 17 and 20).

- (20) a.  $\left[ \text{IIMP} : \text{waroori}^{\text{H}} \Rightarrow \text{NPST} : \left\{ \begin{array}{l} \text{warojui}^{\text{H}} \\ * \text{waroojui}^{\text{H}} \end{array} \right\} \right]$  ‘to laugh’  
 b.  $\left[ \text{IIMP} : \text{koori}^{\text{H}} \Rightarrow \text{NPST} : \left\{ \begin{array}{l} * \text{kojui}^{\text{H}} \\ \text{koojui}^{\text{H}} \end{array} \right\} \right]$  ‘to buy’

Moreover, long vowels cannot appear in the final syllable of converb forms for lexemes with a non-alternating part ending with either *a*, *u* or *o*, due to the ban on super-heavy syllables (4.2.2). Lexically long vowels shorten, and the distinction between short and long vowels is thus neutralised (21).

- (21) a.  $\left[ \text{CVB} : \text{waroi}^{\text{H}} \Rightarrow \text{NPST} : \left\{ \begin{array}{l} \text{warojui}^{\text{H}} \\ * \text{waroojui}^{\text{H}} \end{array} \right\} \right]$  ‘to laugh’  
 b.  $\left[ \text{CVB} : \text{koi}^{\text{H}} (* \text{kooi}^{\text{H}}) \Rightarrow \text{NPST} : \left\{ \begin{array}{l} * \text{kojui}^{\text{H}} \\ \text{koojui}^{\text{H}} \end{array} \right\} \right]$  ‘to buy’

The impact of vowel length differs between the different implications. For example, the values of  $H(\text{NPST} \mid \text{SEQ})$ ,  $H(\text{CVB} \mid \text{SEQ})$  and  $H(\text{IIMP} \mid \text{SEQ})$  are similar when vowel length is ignored, i.e. respectively 0.428, 0.410, and 0.408 (Table 17). However, they are much different when vowel length is taken into account, i.e. respectively 1.381, 1.075, and 0.410 (Table 17). These differences depend on whether there is vowel length neutralisation in the predicted form too.

For almost all lexemes, vowel length neutralisation occurs in the indirect imperative form, whose conditional entropy  $H(\text{IIMP} \mid \text{SEQ})$  is thus the lowest, and since there is neutralisation in both the imperative and the sequential, there is no uncertainty about vowel length between the two (22). On the other hand, neutralisation occurs in the converb only for lexemes with a vocalic segmental alternation

(23a), while in other cases neutralisation in the sequential prevents predicting the length of the vowel in the converb (23b). Finally, predicting the non-past from the sequential is hindered by the absence of vowel length neutralisation in the non-past (24).

- (22) a. [SEQ :  $w^?iizji^H \Rightarrow$  IIMP :  $w^?igji^H$ ] ‘to swim’  
 b. [SEQ :  $tuuti^{LH} \Rightarrow$  IIMP :  $tuuri^{LH}$ ] ‘to take’
- (23) a. [SEQ :  $tuuti^{LH} \Rightarrow$  CVB :  $tui^{LH}$ ] ‘to take’  
 b. [SEQ :  $w^?iizji^H \Rightarrow$  CVB :  $\left\{ \begin{array}{l} w^?igji^H \\ *w^?igji^H \end{array} \right\}$ ] ‘to swim’
- (24) a. [SEQ :  $w^?iizji^H \Rightarrow$  NPST :  $\left\{ \begin{array}{l} w^?igjui^H \\ *w^?igjui^H \end{array} \right\}$ ] ‘to swim’  
 b. [SEQ :  $tuuti^{LH} \Rightarrow$  NPST :  $\left\{ \begin{array}{l} *tuujui^{LH} \\ tujui^{LH} \end{array} \right\}$ ] ‘to take’

### 5.3.3 Tone alternations

The last source of uncertainty comes from the neutralisation of tone patterns. It is however less frequent, and its impact on entropy is thus restricted (Table 19).

First, in Asama, a word cannot begin with a high tone and end with a light syllable bearing a high tone, so that there is no  $*XCV^H$  form with a high tone on every syllable, and  $*CV^H$  forms are also banned due to a minimality constraint. If a word begins with a high tone, then a sequence of final light syllables is automatically low,<sup>9</sup> so that the opposition between the H and HL patterns is neutralised for such forms. Consequently, the HL pattern of a CVVCV form can thus alternate with either H or HL in other paradigm cells (25).

- (25) [SEQ :  $aadi^{H=HL} \Rightarrow$  PROG.NEG :  $\left\{ \begin{array}{l} *aaduran^H \\ aaduran^{HL} \end{array} \right\}$ ] ‘to knit’

The uncertainty caused by tone alternations also contributes to the difference in the values of  $H(\text{PROG} \mid \text{NPST})$  (Table 17) when taking into account suprasegmental alternations (1.247) or not (0.409). This is due to the fact that for lexemes where  $R$  has the structure (C)VVCV, there is an overlap between suprasegmental classes I1 and II2 (26).

- (26) a. [NPST :  $asibjui^H \Rightarrow$  PROG :  $\left\{ \begin{array}{l} *asiduuran^H \\ asiduuran^{LH} \end{array} \right\}$ ] ‘to play’

### 5.3.4 Summary

The impact of each of the three sources of uncertainty, i.e. segmental, vowel length, and tone alternations, is summarised in Table 21.

## 5.4 Binary implications

Table 15 shows that no paradigm cell has a constant conditional entropy of 0 when used as a predictor in unary implications. No single paradigm cell thus constitutes a principal part (Stump & Finkel 2013: p. 11) sufficient to predict all the other forms of a verb’s paradigm. We must thus next turn to binary, i.e. two-to-one, implications.

<sup>9</sup>Except for some recent loanwords.

Table 21: Summary of the sources of uncertainty (V = vowel length, T = tone, S = segment); a dash in a cell indicates that the conditional entropy is null and that there is thus no source of uncertainty to identify

$H(\text{Col}   \text{Row})$	NPST	CVB	IIMP	SEQ	PROG.NEG
NPST		—	—	S	T, S
CVB	V		V	V, S	V, T, S
IIMP	V, T	V, T		V, S	V, T, S
SEQ	V, T, S	V, T, S	V, S		V, T
PROG.NEG	V, S	V, S	V, S	—	

Table 22: Binary implication system

$H(\text{Col}   \text{Row})$	NPST	CVB	IIMP	SEQ	PROG.NEG	Predictiveness
NPST		0	0	0.200	0.228	0.107
{NPST, CVB}	0		0	0.200	0.228	0.107
{NPST, IIMP}	0	0		0.200	0.228	0.107
{NPST, SEQ}	0	0	0		0.023	0.006
{NPST, PROG.NEG}	0	0	0	0		0

We started with the non-past since it is the best unary predictor, i.e. it has the lowest mean conditional entropy when acting as predictor in unary implications (Table 15). We combined the non-past with every other cell, and for every such pair of predictors we computed the corresponding conditional entropy of every cell not included in the pair. The results are given in Table 22, together with the value for unary implications predicted from the non-past for comparison.

Table 22 shows that only the predictor pair {NPST, PROG.NEG} reduces the entropy to zero and is thus able to fully predict the rest of the paradigm. This pair thus constitutes a set of principal parts of Asama’s inflectional system.<sup>10</sup> This naturally follows from the fact that the uncertainty in the realisation of the non-past is due to segmental alternations (§4.2.1), and that segmental alternations are concentrated in zone T<sub>5</sub> (Table 6), to which the negative progressive non-past belongs.

The verb *sjabijui*<sup>H</sup> ‘to rust’ illustrates the difference between unary (27–28) and binary (29) implications. In both (27) and (28), there is uncertainty in the implication and several realisations are possible for some cells, but which cells are affected depends on the predictor. There is uncertainty in the realisation of the sequential form for the non-past (27c), and in the realisation of the converb (28a) and indirect imperative (28b) form for the progressive. The intersection of the two sets of possible realisations for a given cell in (27) and (28) is always a single form. The {NPST, PROG.NEG} pair can thus fully predict the whole paradigm (29).

- (27) a. [NPST : *sjabijui*<sup>H</sup> ⇒ CVB : *sjabiri*<sup>H</sup>]  
b. [NPST : *sjabijui*<sup>H</sup> ⇒ IIMP : *sjabiiri*<sup>H</sup>]  
c.  $\left[ \text{NPST : } sjabijui^H \Rightarrow \text{SEQ : } \left\{ \begin{array}{l} *sjabiiti^H \\ sjabiicji^H \\ *sjabicji^H \end{array} \right\} \right]$

<sup>10</sup>Recall however that these cells were arbitrarily chosen as representative of full interpredictability zones, so that other actual pairs of cells also constitute valid principal parts.



- (28) a.  $\left[ \text{PROG.NEG} : \text{sjabicjuuran}^H \Rightarrow \text{CVB} : \left\{ \begin{array}{l} \text{sjabiri}^H \\ * \text{sjabikii}^H \end{array} \right\} \right]$   
 b.  $\left[ \text{PROG.NEG} : \text{sjabicjuuran}^H \Rightarrow \text{IIMP} : \left\{ \begin{array}{l} \text{sjabiiri}^H \\ * \text{sjabiiki}^H \end{array} \right\} \right]$   
 c.  $[\text{PROG.NEG} : \text{sjabicjuuran}^H \Rightarrow \text{SEQ} : \text{sjabii}^H]$
- (29) a.  $[\{\text{NPST} : \text{sjabijui}^H, \text{PROG.NEG} : \text{sjabicjuuran}^H\} \Rightarrow \text{CVB} : \text{sjabiri}^H]$   
 b.  $[\{\text{NPST} : \text{sjabijui}^H, \text{PROG.NEG} : \text{sjabicjuuran}^H\} \Rightarrow \text{IIMP} : \text{sjabiiri}^H]$   
 c.  $[\{\text{NPST} : \text{sjabijui}^H, \text{PROG.NEG} : \text{sjabicjuuran}^H\} \Rightarrow \text{SEQ} : \text{sjabii}^H]$

Adding the knowledge of the realisation of the converb and the indirect imperative forms to that of the non-past form has no impact on uncertainty, which is also expected given that both the converb and the indirect imperative forms are entirely predictable from the non-past one (see Table 15).

## 6 Conclusions

Our study shows how a Word-and-Paradigm approach coupled with a computational implementation provides interesting qualitative and quantitative insights into the inflectional verb morphology of an endangered and underdescribed language such as Asama. Our approach with conditional entropy measures is able to identify and quantify the different sources of uncertainty in implicative relations, i.e. when trying to predict the realisation of inflectional form from that of another form. In the case of Asama, the uncertainty due to suprasegmental alternations is lower than that due to segmental alternations. Applying entropy measures to not only unary but also  $n$ -ary implications also allows for the identification of the principal parts of the system. For instance, in Asama, conditional entropy reaches zero for all forms when taking the non-past and the negative progressive as predictors, which shows that this set of two cells constitute the principal parts of the system. Though our results highlight the benefits of factoring different types of alternations, e.g. segmental vs. suprasegmental, it is nevertheless necessary to consider alternation patterns based on whole forms rather than isolated exponents due to the existence of fine-grained interdependencies among different factors. In Asama, not only are segmental alternations constrained by the phonotactic properties of the non-alternating part, but there are also interactions between vowel length and tone alternations, which forbids the precise estimation of their respective contribution to uncertainty.

Our work could perhaps be further improved by considering in more detail the phonotactic properties of the non-alternating part of inflected forms, and in particular the relative frequency of vowels which are lexically specified as long. Since automatic vowel lengthening accounts for a large part of the overall entropy of the system, adding such information could lead to different estimations, though the magnitude and the significance of these differences remain to be determined. The inclusion of irregular verbs and their impact on entropy measurements could also be considered, but here also a trade-off between the costs of adding more information to the model and the resulting gains needs to be found. Overfitting the model with additional parameters for every individual lexeme cannot be recommended. Since our corpus contains only 11 irregular verbs, we do not believe the entropy results and our conclusions would be much different even if we included those.

Formal and implemented models based on standard computational tools can greatly benefit the description of endangered languages and help cope with the insufficiency of paper-and-pencil linguistics, which would open up new opportunities for quantitative typological and theoretical studies.



**Acknowledgements** The authors would like to thank their main informant, Okamura Takahiro, who has for several decades devoted his time to the documentation and preservation of his native language. We also thank the two anonymous reviewers as well as the editor for their useful comments.

## References

- Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–81. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199547548.003.0003>.
- Ackerman, Farrell & Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89(3). 429–464. <https://doi.org/10.1353/lan.2013.0054>.
- Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge: The MIT Press.
- Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2017. *Morphological complexity*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316343074>.
- Bauer, Laurie. 2016. Classical morphemics: Assumptions, extensions, and alternatives. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge handbook of morphology*, 331–355. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781139814720.013>.
- Beniamine, Sacha. 2018. *Classifications flexionnelles: Étude quantitative des structures de paradigmes*. Paris: Université Sorbonne Paris Cité dissertation. <https://tel.archives-ouvertes.fr/tel-01840448>.
- Beniamine, Sacha, Olivier Bonami & Ana R. Luís. 2021. The fine implicative structure of European Portuguese conjugation. *Isogloss* 7. 1–35. <https://doi.org/10.5565/rev/isogloss.109>.
- Blevins, James P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199593545.001.0001>.
- Blevins, James P. & Juliette Blevins (eds.). 2009. *Analogy in grammar: Form and acquisition*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199547548.001.0001>.
- Bonami, Olivier. 2014. *La structure fine des paradigmes de flexion: Études de morphologie descriptive, théorique et formelle*. Paris: Université Paris Diderot habilitation dissertation. <https://tel.archives-ouvertes.fr/tel-01531860>.
- Bonami, Olivier & Sacha Beniamine. 2016. Joint predictiveness in inflectional paradigms. *Word Structure* 9(2). 156–182. <https://doi.org/10.3366/word.2016.0092>.
- Brown, Dunstan & Andrew Hippisley. 2012. *Network Morphology: A defaults-based theory of word structure*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511794346>.
- Chen, Emily, Hyunji Hayley Park & Lane Schwartz. 2020. Improved finite-state morphological analysis for St. Lawrence Island Yupik using Paradigm Function Morphology. In *Proceedings of The 12th Language Resources and Evaluation Conference*, 2676–2684. Marseille: European Language Resources Association. <https://www.aclweb.org/anthology/2020.lrec-1.326>.
- Corbett, Greville G. 2009. Canonical inflectional classes. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, 1–11. Somerville: Cascadilla Proceedings Project.
- Crysmann, Berthold. 2016. Representing morphological tone in a computational grammar of Hausa. *Journal of Language Modelling* 3(2). 463–512. <https://doi.org/10.15398/jlm.v3i2.126>.
- Finkel, Raphael & Gregory Stump. 2007. Principal parts and morphological typology. *Morphology* 17. 39–75. <https://doi.org/10.1007/s11525-007-9115-9>.

- Finkel, Raphael & Gregory Stump. 2009. Principal parts and degrees of paradigmatic transparency. In James P. Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 13–53. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199547548.003.0002>.
- Guzmán Naranjo, Matías. 2020. Analogy, complexity and predictability in the Russian nominal inflection system. *Morphology* 30(3). 219–262. <https://doi.org/10.1007/s11525-020-09367-1>.
- Harrigan, Atticus G. et al. 2017. Learning from the computational modelling of Plains Cree verbs. *Morphology* 27(4). 565–598. <https://doi.org/10.1007/s11525-017-9315-x>.
- Hayata, Teruhiro. 1995. Tokunoshima Amagi-chō Matsubara to Asama no hōgen no meishi akusento taikai: Tango seichō hōgen no rei to shite. *Ryūkyū no Hōgen* 18/19. 132–144. <https://doi.org/10.15002/00012659>.
- Heinrich, Patrick, Shinsho Miyara & Michinori Shimoji (eds.). 2015. *Handbook of the Ryukyuan languages: History, structure, and use*. Berlin: De Gruyter. <https://doi.org/10.1515/9781614511151>.
- Hockett, Charles F. 1987. *Refurbishing our foundations: Elementary linguistics from an advanced point of view*. Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.56>.
- Hulden, Mans. 2009. Foma: A finite-state compiler and library. In *Proceedings of the Demonstrations Session at EACL 2009*, 29–32. <https://www.aclweb.org/anthology/E09-2008>.
- Jacques, Guillaume et al. 2012. An overview of Khaling verbal morphology. *Language and Linguistics* 13(6). 1095–1170. <https://hal.science/hal-01361190>.
- Karttunen, Lauri. 2006. The insufficiency of paper-and-pencil linguistics: The case of Finnish prosody. In Miriam Butt, Mary Dalrymple & Tracy Holloway King (eds.), *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*, 287–300. Stanford: Center for the Study of Language and Information Publications. <https://roa.rutgers.edu/article/view/828>.
- Lawrence, Wayne. 2015. Tokunoshima Asama hōgen no meishi onchō ni tsuite nidai: Kōki onchō to shishō gokei no onchō. *Ryūkyū no Hōgen* 40. 1–17. <https://doi.org/10.15002/00013673>.
- Lévêque, Dimitri. 2021. *Grammaire de l'asama*. Paris: Inalco dissertation.
- van der Lubbe, Gijs & Akiko Tokunaga. 2015. Okinoerabu grammar. In Patrick Heinrich, Shinsho Miyara & Michinori Shimoji (eds.), *Handbook of the Ryukyuan languages: History, structure, and use*, 345–378. Berlin: De Gruyter. <https://doi.org/10.1515/9781614511151.345>.
- Matthews, Peter H. 1972. *Morphology*. 1st edn. Cambridge: Cambridge University Press.
- Niinaga, Yuto. 2014. *A grammar of Yuwan, a northern Ryukyuan language*. Tokyo: University of Tokyo dissertation. <https://doi.org/10.15083/00007642>.
- Niinaga, Yuto. 2015. Amami grammar. In Patrick Heinrich, Shinsho Miyara & Michinori Shimoji (eds.), *Handbook of the Ryukyuan languages: History, structure, and use*, 323–344. Berlin: De Gruyter. <https://doi.org/10.1515/9781614511151.323>.
- Okamura, Takahiro et al. (eds.). 2009. *Tokunoshima hōgen nisen bun jiten*. revised. Matsumoto: Tokunoshima Hōgen no Kai.
- Parker, Jeff & Andrea D. Sims. 2020. Irregularity, paradigmatic layers, and the complexity of inflection class systems: A study of Russian nouns. In Peter Arkadiev & Francesco Gardani (eds.), *The complexities of morphology*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198861287.001.0001>.
- Pellard, Thomas. 2015. The linguistic archeology of the Ryukyu Islands. In Patrick Heinrich, Shinsho Miyara & Michinori Shimoji (eds.), *Handbook of the Ryukyuan languages: History, structure, and use*, 13–37. Berlin: De Gruyter. <https://doi.org/10.1515/9781614511151.13>.
- Pellard, Thomas & Masahiro Yamada. 2017. Verb morphology and conjugation classes in Dunan (Yonaguni). In Ferenc Kiefer, James P. Blevins & Huba Bartos (eds.), *Perspectives on morphological organization*, 31–49. Leiden: Brill. [https://doi.org/10.1163/9789004342934\\_004](https://doi.org/10.1163/9789004342934_004).

- Robins, R. H. 1959. In defence of WP. *Transactions of the Philological Society* 58(1). 116–144. <https://doi.org/10.1111/j.1467-968X.1959.tb00301.x>.
- Shannon, Claude E. 1948a. A mathematical theory of communication. *The Bell System Technical Journal* 27(3). 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- Shannon, Claude E. 1948b. A mathematical theory of communication. (Concluded from July 1948 issue). *The Bell System Technical Journal* 27(4). 623–656. <https://doi.org/10.1002/j.1538-7305.1948.tb00917.x>.
- Shimoji, Michinori. 2017. *A grammar of Irabu: A Southern Ryukyuan language*. Fukuoka: Kyushu University Press.
- Snoek, Conor et al. 2014. Modeling the noun morphology of Plains Cree. In *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, 34–42. <https://doi.org/10.3115/v1/W14-2205>.
- Stump, Gregory. 2016. *Inflectional paradigms*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9781316105290>.
- Stump, Gregory & Raphael. A. Finkel. 2013. *Morphological typology: From word to paradigm*. Cambridge: Cambridge University Press.
- Uwano, Zendō. 1977. Tokunoshima Asama hōgen no akusento (1). In Iwate Kokugo Gakkai Ronshū Kankōkai (ed.), *Kokugogaku ronshū: Komatsushiro Yūichi kyōju taishoku, Shima Minoru kyōju taikan kinen*, 188–220 (1–33). Morioka: Iwate Kokugo Gakkai Ronshū Kankōkai.
- Uwano, Zendō. 2001. Tokunoshima Asama hōgen no katsuyōkei akusento shiryō. *Ryūkyū no Hōgen* 25. 1–61. <https://doi.org/10.15002/00012560>.
- Uwano, Zendō. 2016. Tokunoshima Asama hōgen no meishi akusento taikai. In Yukinori Takubo, John Whitman & Tatsuya Hirako (eds.), *Ryūkyū shogo to Kodai Nihongo: Nichiryū sogo no saiken ni mukete*, 209–234. Tokyo: Kuroshio Shuppan.
- Wilmoth, Sasha & John Mansfield. 2021. Inflectional predictability and prosodic morphology in Pitjantjatjara and Yankunytjatjara. *Morphology* 31(4). 355–381. <https://doi.org/10.1007/s11525-021-09380-y>.
- Wurzel, Wolfgang Ulrich. 1989. *Inflectional morphology and naturalness*. Dordrecht: Kluwer.
- Yokoyama (Tokunaga), Akiko. 2017. *Ryūkyū Okinoerabu-jima Kunigami hōgen no bunpō*. Kunitachi: Hitotsubashi University dissertation. <https://doi.org/10.15057/30419>.