Description of verbal morphology of Asama: a realizational and implemented approach

Dimitri Lévêque — INalCO - CNRS - EHESS, CRLAO
Thomas Pellard — CNRS - INaLCO - EHESS, CRLAO

In this presentation, we will give an overview of the verbal morphology of Asama, an endangered Japonic language, which includes non-concatenative phenomena implying mostly tone and vowel length. We will introduce the theoretical and practical framework we have used for this study, explain why we chose it and how we used it in order to analyze the verbal morphology of Asama, and conclude by showing how the programs implemented can be used to obtain more quantitative results by computing paradigm cells interpredictability and Shannon entropy.

1 Theoretical and practical framework

Asama is a Japonic language spoken in the village of the same name located in the island of Tokunoshima, Japan. While Japanese and Japonic languages are usually known for their relatively simple agglutinative verbal morphology, it is not the case for all of them, and some like Dunan (Pellard & Yamada, 2017) or Asama (Lévêque, 2017) show a more complex morphology. Asama's verbal morphology indeed displays several phenomena of stem alternation (see Table 1 for an illustration) and suffix allomorphy that are not found elsewhere in Japonic languages, and that make it far less easy to describe with a classical morpheme-based framework (Selkirk, 1982; Halle & Marantz, 1993). All the data used in this study (fifty complete and several hundred partial verbal paradigms) are taken from Okamura et al. (2009), Uwano (2001) and one of the author's own data collected during fieldwork in 2018 and 2019.

Table 1: Stem alternation in Asama

Form	TUBJUI "to fly"	KOORAKJUI "to dry"	MˀAARɨJUI "to be born"
NPST	tubj- (H)	koorakj- (H)	mˀaarɨj- (H)
PROH	tubj- (H)	koorakj- (H)	mˀaarɨj- (H) / mˀaarɨ- (H)
NEG	tub- (H)	koorak- (H)	mˀaarɨr- (H)
DES	tub- (H)	koorak- (H)	mˀaarɨ- (H)
PST	tud- (H)	kooracj- (H)	mˀaarɨt- (H)
SEQ	tud- (H)	kooraacj- (H)	mˀaarɨɨt- (H)
PROG	tud- (LH)	kooracj- (H) / kooraacj	mˀaarɨt- (H) / mˀaarɨɨt- (H)

The verbal morphology was fully implemented with finite-state transducers (Beesley & Karttunen, 2003) and with the foma software (Hulden, 1999).

Asama's verbal morphology has many features that make it far from being canonical, in the sense of Corbett (2009). Two of those features are listed above, and many of those phenomena involve tone and vowel length, two aspects that deserve a special attention in Asama verbal morphology.

- Overabundance: one case of a given paradigm has more than one surface form (ex. 1).

(1)

Meaning	{PROG PST} Var. 1	{PROG PST} Var. 2
”to begin”	hazi**m~~i~~i~~i~~tu**taN (H)	hazi**m~~i~~tuu**taN (H)
”to be born”	maar**~~i~~i~~i~~tu**taN (H)	maar**~~i~~tuu**taN (H)

- Multi-functional exponent: a single form can express more than one morphosyntactic property (ex. 2).

(2)

Meaning	{IMP.INDIR}	{SEQ}
”to read”	juum-**~~i~~** (LH)	juud-**~~i~~** (LH)
”to meet”	oor-**~~i~~** (H)	oot-**~~i~~** (H)

That is why for a great majority of inflectional forms, it is not possible to use a morpheme based analysis, since in many cases a given morpheme cannot be related to a fixed morphosyntactic property or set of properties. On the contrary, a Word-and-Paradigm approach (Hockett, 1954) such as Paradigm Function Morphology (PFM), a framework that has been first developed by Stump (2001), is more promising to account for inflectional morphology of Asama. In this study, we thus follow the six core assumptions posited by Bonami (2014) and Bonami & Stump (2016) to describe PFM. In particular, we base on the assumption of the definition of a language inflectional morphology as a paradigm function in order to analyze and compute verbal morphology.

2 Analyses and results

In order to describe verbal morphology in Asama, we use two different sets of inflectional classes. The first set is meant to take into account the segmental stem alternation alone, as it can be seen in the Table 2 below, that shows a partial description of verbal morphology. In this table, X, Y and Z are purely morphomic categories (Aronoff, 1994), X and Y being subdivised in sub-categories for only a few inflectional classes.

Table 2: First set of inflectional classes

Inflectional class	X_1	X_2	X_3	Y_1	Y_2	Z
m ~ d	mj	*NA*	*NA*	m	m	d
n ~ n	nj	*NA*	*NA*	n	nj	zj
k ~ c	kj	*NA*	*NA*	k	k	cj
k ~ z	kj	*NA*	*NA*	k	k	zj
s ~ s	sj	*NA*	*NA*	s	sj	sj
t ~ cc	cj	*NA*	*NA*	t	t	ccj
i ~ r ~ t	j	-	-	r	r	t
{a,o,u} ~ r ~ tt	j	-	-	-	r	tt
{s,z}~~i~~ ~ r ~ cj	j	*NA*	-	r	r	cj
{a,o,u} ~ r ~ cj	j	-	-	-	r	cj

The second set of inflectional classes accounts for the autosegmental part of stem alternation, implying mainly tone and vowel length. Table 3 shows how it is possible to represent those two classes with a single set of rules (”R” means ”Root”) and can be read as an explicitation of the realisational rules that lays at the core of the Paradigm Function Morphology.

Table 3: Second set of inflectional classes

Classe	{IMP.DIR}	{SEQ}	{PROG NEG}	{PROG NEG NPST}
I_1	R:Y_2i (H)	R:Zi (H)	RZui (H)	RZuuraN (H)
I_2	R:Y_2i (H)	R:Zi (H)	R:Zui (H) / RZui (H)	R:ZuraN (H) / RZuuraN (H)
II_1	RY_2ii (H)	RZii (H)	RZui (LHL)	RZuuraN (LHL)
II_2	R:Y_2i (H)	R:Zi (H)	RZui (LHL)	RZuuraN (LHL)
III	R:Y_2i (HL)	R:Zi (HL)	R:Zui (HL)	R:ZuraN (HL)
IV	R:Y_2i (LH)	R:Zi (LH)	R:Zui (LHL)	R:ZuraN (LHL)

The verbal lexeme AMJUI "to knit", which is associated to the inflectional classes "m∼d" and "III", can be used to illustrate this approach. To obtain the form of this lexeme, associated for instance with the following set of morphosyntactic properties {prog neg npst}, the following realizational rules apply:

- Rules of Stem Choice: $RSC_{md,III}(\langle$AMJUI,σ{prog neg npst}$\rangle) = \langle$aad,$\sigma\rangle$
- Rules of Exponence:
  - Block 1: $RE1_{md,III}(\langle$aad,σ{prog neg npst}$\rangle) = \langle$aadu,$\sigma\rangle$
  - Block 2: $RE2_{md,III}(\langle$aadu,σ{prog neg npst}$\rangle) = \langle$aaduran,$\sigma\rangle$
  - Block 3: $RE3_{md,III}(\langle$aaduran,σ{prog neg npst}$\rangle) = \langle$aaduran (HL),$\sigma\rangle$

Those rules correspond to what is found in the cell in grey color in Table 3. The implementation of all the transducers describing the complete verbal morphology are based on those two tables.

In the presentation, we first show with the help of detailed examples that the model outlined above is both an elegant and efficient way to describe the verbal morphology of the Asama language. It is moreover in line with the primary goal of this study, that is to say the description of an under-documented and undescribed language such as Asama, especially since it helps automating the gloss.

In the second part of the presentation, we present some quantitative results obtained from the transducers. Starting from the notion of paradigmatic structure, implicative morphology (Wurzel, 1989; Albright, 2002; Bonami, 2014) and interpredictability between the cells of paradigms, we have computed, as it has been done for example in Bonami & Luis (2015) or Pellard & Yamada (2017), interpredictability and conditional entropy calculations (Shannon, 1948; Blevins, 2013; Ackerman et al., 2009; Ackerman & Malouf, 2013), that give further strength to the qualitative explanations. All the calculations are made on the base of a distillation of the paradigm (Stump & Finkel, 2013, 42), that results in a sub-paradigm of five cells, each of them representing one area of full interpredictability of the full paradigm.

Entropy results are first used to highlight the main sources of uncertainty, and thus confirm what has been found in the qualitative analysis. Uncertainty mainly lies in neutralisations of oppositions based on tone and vowel length, and also in unpredictable segmental alternations. We also take advantage of the programs to attribute a weight to the uncertainty that can be attributed to segmental alternations and the uncertainty that can be attributed to suprasegmental alternations.

Finally, entropy calculations based on binary implications show that the best candidates to the status of principal parts of the verbal morphology (Stump & Finkel, 2013) are the nonpast and progressive forms.

References

Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins & J. Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–81. Oxford: Oxford University Press.

Ackerman, Farrell & Robert Malouf. 2013. Morphological Organization: The Low Conditional Entropy Conjecture. *Language* 89. 429–464.

Albright, Adam C. 2002. *The Identification of Bases in Morphological Paradigms*. Los Angeles: University of California dissertation.

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes*. Cambridge: MIT Press.

Beesley, Kenneth R. & Lauri Karttunen. 2003. *Finite State Morphology*. Stanford, CA: CSLI Publications.

Blevins, James P. 2013. The information-theoretic turn. *Psihologija* 46(3). 355–375.

Bonami, Olivier. 2014. *The fine structure of inflection paradigms*. Paris: Université Paris 7 - Denis Diderot. Habilitation à diriger des recherches.

Bonami, Olivier & Ana R. Luis. 2015. Sur la morphologie implicative dans la conjugaison du portugais : une étude quantitative. In *Mémoires de la société de linguistique de paris*, vol. 22, 111–151. Paris.

Bonami, Olivier & Gregory T. Stump. 2016. *Cambridge Handbook of Morphology* chap. Paradigm Function Morphology, 449–481. Cambridge: Cambridge University Press.

Corbett, Greville G. 2009. Canonical inflectional classes. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, Somerville: Cascadilla Proceedings.

Halle, Morris & Alex Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from building 20*, 111–176. Cambridge, MA: MIT Press.

Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10. 210–231.

Hulden, Mans. 1999. Foma: a finite-state compiler and library. *Association for Computational Linguistics* 29–32.

Lévêque, Dimitri. 2017. *Description implémentée de la phonologie et de la morphologie verbale du dialecte des langues ryukyu parlé à Asama, Tokunoshima*. Paris: Paris III Sorbonne-Nouvelle / INaLCO M.A. Thesis.

Okamura, Takahiro, Motoei Sawaki, Yumi Nakajima, Chitsuko Fukushima & Satoru Kikuchi. 2009. *Tokunoshima hōgen nisen bun jiten*. Matsumoto: Shinshū daigaku Jinbun gakubu.

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. Brill. doi:10.1163/9789004342934_004.

Selkirk, Elizabeth. 1982. *The syntax of words*. Cambridge, MA: MIT Press.

Shannon, Claude E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27. 379–423 and 623–656.

Stump, G. & R. Finkel. 2013. *Morphological typology*. Cambridge: Cambridge University Press.

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

Uwano, Zendō. 2001. *Tokunoshima asama hōgen no katsuyōkei akusento shiryō*. Hosei University.

Wurzel, Wolfgang Ulrich. 1989. *Inflectional Morphology and Naturalness*. Dordrecht: Kluwer.