Covert systematicity in a distributionally complex system[1]

MATTHEW BAERMAN

Surrey Morphology Group, University of Surrey

Current thinking on inflection classes views them as organized networks rather than random assemblages of allomorphs (a view that reaches back to the 1980s, with such notions as Wurzel's paradigm structure conditions and Carstairs's paradigm economy). But we still find systems which appear to lack any visible implicative structure. A particularly striking example comes from Võro (a variety of South Estonian). Its system of verbal inflectional suffixes is formally simple but distributionally complex: although there are never more than three allomorphs in competition, nearly two dozen inflectional patterns emerge through rampant cross-classification of the allomorphs. Allomorph choice in one part of the paradigm thus fails to constrain allomorph choice in the rest, so it looks as if the paradigms would have to be memorized en masse. The key to these patterns lies outside the system of suffixation itself, in the more conventional formal complexity of stem alternations and their paradigmatic patterning. The computationally implemented analysis presented here provides a model of inflection in which the implicational network of phonological, morphophonological and morphological conditions on formal realization are unified in a single representation.

1. INTRODUCTION

It is standard practice, both in descriptive works and theoretical analyses, to regard variation in the form of inflectional exponents as dividing the lexicon into different inflection classes. This reflects not just a natural human (or at any rate linguist's) urge to categorize, but follows from real properties of the data. Consider the material in Table 1, which illustrates suffix allomorphy in the verbs of Orokaiva, a language of the Binanderean family (Trans New Guinea Phylum) of Papua New Guinea. The morphosyntactic values shown are each realized by one of two or three distinct suffix allomorphs; the three

---

	Suffixes	'cook'	'sleep'	'eat'
ABRUPT IMPERATIVE	e, a	ag-e	ev-e	ind-a
PUNCTILIAR SEQUENCE	eto, uto, ito	ag-eto	ev-uto	ind-ito
FAR PAST PL	ea, ua, ia	ag-ea	ev-ua	ind-ia
FAR PAST HABITUAL 1SG	itiaetena, atena	ag-itiaetena	ev-itiaetena	ind-atena
FUTURE 1SG	asona, esona	ag-asona	ev-esona	ind-esona
POTENTIAL 1SG	asina, esi, esina	ag-asina	ev-esi	ind-esina
DESIDERATIVE	asi, esi	ag-asi	ev-esi	ind-esi
NEG. IMPERATIVE	aojo, ojo	ag-aojo	ev-ojo	ind-ojo

Table 1
Suffix allomorphy and inflection classes in Orokaiva verbs (Larsen 1977: 11).

verbs 'cook', 'sleep' and 'eat' exemplify three different patterns of their distribution. In theory, the choice of suffix allomorph for each value could vary independently, in which case one could generate up to 288 different patterns by freely combining the various allomorphs. But in fact, the three patterns in Table 1 are the only ones attested. These patterns are thus more than just the sum of their parts – the allomorphs are bound to each other by a tight network of implicational relationships, and this is what we call inflection classes.

Since Wurzel (1984) and Carstairs (1983), the structure of inflection classes has come to play an ever more prominent role in morphological theory. The particular relevance of inflection classes lies in their being an instance of pure morphological structure, in as much as inflection classes, as traditionally and canonically understood, are arbitrary form-based classes, and not exponents of some syntactic or semantic distinction. (If they were we would more likely be inclined to give them a label based on their function, e.g. TRANSITIVE VERBS or ANIMATE NOUNS, though local traditions differ in this respect.) Perhaps of most interest are suggestions that there is a shared set of organizational principles behind inflection classes across what are unrelated and, at least superficially, profoundly different systems.

At the core of these suggestions is the idea that in spite of their potential for great formal complexity, inflection classes tend towards an efficient organization that allows the generation of paradigms on the basis of relatively little prior information. Carstairs's (1983) Paradigm Economy Principle (PEP) was an early proposal; it states that the number of inflection classes in a language will be far closer to the minimum that is logically possible rather than the maximum, given a particular set of allomorphs (see Müller 2007 for a thorough overview and assessment of this and subsequent proposals). The Orokaiva paradigm above illustrates this, displaying the logical minimum of three classes, which is far from the logical maximum of

2

288 classes. The practical consequence of this is that knowledge of the inflection class membership of a lexeme can be reduced to knowledge of a single form, which can be construed as a principal part, namely a reference form from which the rest of the paradigm can be deduced (Finkel & Stump 2007). To the extent that it is true it is striking, because it is not as if the inflection classes here were monolithic entities; thus 'cook' and 'sleep' classes sometimes pattern together (e.g. in the abrupt imperative), sometimes the 'sleep' and 'eat' classes do (e.g. in the desiderative). Even these apparently well-behaved paradigms contain the seeds of complexity. Indeed, it has become clear that a principle such as the PEP is only a first approximation. While no system has yet been found as wildly complex as the hypothesized pseudo-Orokaiva with its 19 allomorphs generating 288 inflection classes, there are languages where the degree of cross-classification within the inflectional classes steers them far enough away from the logical minimum that one wonders if any principles whatsoever govern their organization. This paper focusses on a particularly striking example, the inflectional suffixes of verbs in Võro.

1.1 *Võro*

Võro is a variety of South Estonian, classed by some as a separate language, which is separated from North Estonian varieties (the basis of contemporary Standard Estonian) by a number of isoglosses which are as old as the break-up of Proto-Balto-Finnic into Finnish, Estonian, Livonian etc. (Viitso 2003). Iva (2007) has provided a comprehensive description of its inflectional morphology, which is the basis for the account here. In the portion of the system shown in Table 2, there are exactly the same number of morphosyntactic values (eight) and allomorphs (nineteen) as in the Orokaiva example. Table 2 also provides some exemplary forms.[2] As a first clue into the nature of this

---

[2] The orthography is largely the same as that of standard Estonian. The main points to note are:

õ = /ɤ/
d/t, g/k = lenis/fortis stop
`σ = 'quantity 3' (overlong) syllable (symbol used only in linguistics works)
q = /ʔ/ (not found in Standard Estonian)
t̬ = /tʲ/ (not found in Standard Estonian)

Note that Iva (2007) only marks quantity 3 syllables where this is not otherwise deduceable from the orthography. In order to spare the reader additional computation, I have added the notation '`' in these contexts as well.

Morphosyntactic feature values are abbreviated as follows: ABL = ablative, ACC = accusative, COND = conditional, DAT = dative, EVID = evidential, GEN = genitive, GER = gerund, ILL = illative, IMP = imperative, IMPERS = impersonal, INF = infinitive, INS = instrumental, JUSS = jussive, LOC = locative, NOM = nominative, PL = plural, PRS = present, PST = past, PTCP = participle, SG = singular, SUP = supine.

3

	Suffixes	'neigh'	'get used to'	'stay'	'call'
PRS 3SG	Ø, s	`hirn	harinõ-s	`jää-s	helise-s
PST I	V, si	hirni	hari-si	`jäi	helisi
INF	Aq, daq, taq	`hirnu-q	harinõ-daq	jäi-äq	helis-täq
GER	En, dEn, tEn	`hirnu-n	harinõ-dõn	jäi-en	helis-ten
JUSS	kuq, guq	`hirn-kuq	hari-guq	jää-guq	helis-kuq
IMPERS PRS 3SG	tAs, dAs, As	hirnu-tas	harinõ-das	jäi-äs	helis-täs
IMPERS PTCP PST SG	t, d	hirnu-t	harinõ-t	`jää-d	helise-t
PTCP PRS	va, v	`hirn-va	harinõ-v	`jää-vä	helise-v

Note: A = /a, ä/, E = /õ, e/, depending on vowel harmony; V is a vocalic suffix subject to further conditions (see Table 7).

Table 2
Suffix allomorphy in Võro verbs.

system, note that although no morphosyntactic value is realized by more than three allomorphs, we need at least four paradigms to illustrate the whole repertoire. In fact, the ability of the system to generate different patterns is much greater than this. When all the attested combinations of the suffixes are spelled out, 23 types emerge, as shown in Table 3.[3] While this is certainly closer to the logical minimum than the logical maximum, its distance from the minimum is still considerable.

Clearly, inflection class assignment here cannot be reduced to knowledge of a single form, as no single form allows for more than a rough triage of the classes. Viewed in these terms, the full paradigm of a lexeme can only be deduced on the basis of multiple forms. This is the insight behind the invocation of principal parts, a traditional notion that has enjoyed a resurgence of attention in recent years (Blevins 2006; Finkel & Stump 2007, 2009). There are various ways of evaluating the structure of principal parts, though by any metric what we see in Table 3 is complex. A straightforward way of looking at this question is through the traditional practice of positing a fixed list of principal parts for the entire system ('static principal parts' per Finkel & Stump 2007). For example, the four principal parts of the Latin verb are the ISG present active, the infinitive, the ISG perfect active, and the supine, which the student is enjoined to remember for each new verb in order to know how to inflect it. On this approach, all eight forms listed in Table 3 would have to be principal parts (as established by Finkel & Stump's *Analyzing Principal*

---

[3] This differs slightly from Iva's (2007) classification, largely in the conflation of distinctions which are treated in this paper as predictable. On the other hand, it factors in the lack of complete interpredictability between the infinitive and gerund (Iva 2007: 107), which is not a parameter in Iva's system. The mapping between Iva's exemplary paradigms and the types here is given in Appendix C.

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
PRS 3SG	Ø	Ø	Ø	Ø	Ø	Ø	s	s	s	s	s	s
PST I	si	V	V	V	V	V	si	si	si	si	si	V
INF	Aq	Aq	Aq	Aq	Aq	tAq	dAq	dAq	dAq	dAq	tAq	Aq
GER	En	En	En	En	En	tEn	dEn	dEn	tEn	tEn	tEn	En
JUSS	guq	guq	guq	guq	kuq	kuq	guq	guq	guq	kuq	kuq	guq
IMPERS PRS 3SG	tAs	As	dAs	tAs	tAs	tAs	dAs	tAs	tAs	tAs	tAs	As
IMPERS PTCP PST SG	t	d	t	t	t	tt	t	t	t	t	t	d
PTCP PRS	v	vA	v	v	vA	v	v	v	v	v	v	vA
NO. OF LEXEMES	88	15	3	87	689	2	58	127	9	437	13	3

	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX	XXI	XXII	XXIII
PRS 3SG	s	s	s	s	s	s	s	s	s	s	s
PST I	V	V	V	V	V	V	V	V	V	V	V
INF	Aq	Aq	Aq	Aq	Aq	Aq	dAq	dAq	dAq	dAq	tAq
GER	En	En	En	En	En	En	dEn	dEn	dEn	dEn	tEn
JUSS	guq	guq	guq	guq	kuq	kuq	guq	guq	guq	guq	kuq
IMPERS PRS 3SG	dAs	dAs	tAs	tAs	dAs	tAs	dAs	dAs	tAs	tAs	tAs
IMPERS PTCP PST SG	d	t	t	t	t	t	d	t	t	t	t
PTCP PRS	v	v	v	vA	v	vA	v	v	v	vA	v
NO. OF LEXEMES	355	1601	24	5	115	18	257	230	393	6	133

Table 3
The suffixes in Table 2, divided into classes (adapted from Iva 2007: 121f.).

		1st declension	2nd declension	
	Suffixes	'wax'	Stem in -r 'boy'	Other stems 'disciple'
NOM SG	a, Ø, us	cēr-a	puer	discipul-us
ACC SG	am, um	cēr-am	puer-um	discipul-um
GEN SG	ae, ī	cēr-ae	puer-ī	discipul-ī
DAT SG	ae, ō	cēr-ae	puer-ō	discipul-ō
ABL SG	ā, ō	cēr-ā	puer-ō	discipul-ō

Table 4
Conditioned allomorphy within the Latin 2nd declension.

Parts software).[4] This highlights the particular tension created by such a system. On the one hand, there is a highly restricted inventory of forms, which suggests a rule-based system. On the other hand, the distributional complexity of these forms would seem to require lexical storage of a large set of principal parts. But storage, at least as normally construed, involves unusual forms, e.g. *children* as the plural of *child*. What we see in Table 3 are not unusual forms, but rather unusual – or at least varying – distributions of what are otherwise ordinary forms. Storage of 'ordinary' forms is certainly a possibility to be considered, though the convincing cases that have been described restrict this to high frequency items (Baayen et al. 2003). In Võro, too much of the lexicon is involved for this to be a plausible approach. What then does knowledge of such a system actually constitute?

1.2 *Inflectional interactions*

The key to understanding the nature of this system is the other components of the inflectional paradigm, which provide additional information which the rules of suffixation can tap into. Such interactions have been the focus of an increasing body of work (Carstairs 1983, Brown et al. 1996, Cameron-Faulkner & Carstairs-McCarthy 2000, Blevins, Ackerman & Malouf 2009, Stump & Finkel 2010), and it is the purpose of this paper to make an additional contribution to this research programme. A simple example of such an analysis is shown in Table 4, which illustrates the singular forms of three distinct noun paradigms in Latin. Traditionally these three types are subsumed under just two inflection classes: the 'wax' type as 1st declension and the other two as 2nd declension. The difference between the two 2nd declension paradigms is correlated with stem phonology: 2nd declension stems

[4] Available at: http://www.cs.uky.edu/~raphael/linguistics/analyze.html.

6

	Suffixes	Truncating 'Armenian'	Non-truncating 'Georgian'
NOM SG	Ø	armjan-in	gruz-in
ACC/GEN SG	a	armjan-in-a	gruz-in-a
DAT SG	u	armjan-in-u	gruz-in-u
LOC SG	e	armjan-in-e	gruz-in-e
INS SG	om	armjan-in-om	gruz-in-om
NOM PL	e, y	armjan-e	gruz-in-y
ACC/GEN PL	Ø	armjan	gruz-in
DAT PL	am	armjan-am	gruz-in-am
LOC PL	ax	armjan-ax	gruz-in-ax
INS PL	ami	armjan-ami	gruz-in-ami

Table 5
Conditioned allomorphy of Russian ethonyms in *-in*.

ending in /r/ almost exclusively take Ø in the nominative singular, while other stems take *-us*. Thus, the three paradigm types are ascribed to two different sources: lexical specification on the one hand, and stem-phonology–based allomorph assignment on the other. (Note that the latter is still morphological in a sense, as it is subordinate to inflection class: a stem-final /r/ in the 1st declension has no effect on suffix assignment.)

Alternatively, the supplementary information cuing the assignment of allomorphs may itself be paradigmatic, but from a different subsystem, e.g. stem alternation as opposed to affixation Table 5 illustrates an example from Russian. Ethonyms ending in *-in* fall into two types on the basis of their stem alternation pattern. In one type this *-in* is truncated in the plural, in others it is a fixed part of the stem: compare the nominative singular ∼ plural forms *armjanin* ∼ *armjan-e* 'Armenian(s)' with *gruzin* ∼ *gruzin-y* 'Georgian(s)'. This difference in stem alternation correlates with a difference in the nominative plural suffix: the truncating type has *-e* in place of the default suffix *-y*. It is not predictable directly from the stem (mere lack of a stem-final *-in* does not imply *-e*; compare *grubijan* ∼ *grubijan-y* 'rude person ∼ people'), only from the overall pattern of singular ∼ plural alternation. Thus, so long as information about the stem alternation pattern is available, the suffix allomorphy can be derived from it without resorting to dedicated lexical specification.

The striking complexity of the Võro suffix classes, characterized by the almost random distribution of allomorphs, can largely be ascribed to an especially extreme form of conditioned allomorphy. Ultimately, the distributional complexity of the suffix system can be derived from more conventional irregularity of form (in particular, stem alternations), along with the number and variety of conditions that allomorphy is sensitive to.

7

The system of LEXICALLY SPECIFIED inflection classes ends up being fairly simple, though the overall inflectional system it is embedded in is not. There is, however, no easy way to segregate inflection classes in the strict sense from other types of suffix allomorph assignment, because the various elements are interwoven. Nor is there any reason to do so; the aim of the present study is to present a unified model of this complex inflectional system, one which nevertheless allows the contributions of the various components to be tracked.

1.3 *The data*

The basis of this study is Iva's (2007) description of the nominal and verbal inflectional morphology of Võro, in particular his three-way classification of the inflectional paradigm according to (i) suffixation, (ii) stem-final segment alternations, and (iii) stem gradation (a type of stem alternation; see Section 3.2), itself inspired by Viks's (1992) similar treatment of Estonian. I have simplified this system by adopting a somewhat more abstract analysis of the suffixes themselves, largely by ascribing the quality of the past 1st person exponent -*V* to separate morphophonological operations (see Table 7 below), and different segmentation of the impersonal (see Section 2 below). This reduces Iva's 34 classes to the 23 in Table 3.[5] I have also consolidated the various stem alternation types into larger macro-classes, since not all of the distinctions are relevant for suffix assignment. The original contribution of the present study to our understanding of the data is the identification of the implicational relationships between these systems (including the role of prosody, which is not addressed by Iva) outlined in Section 3, leading to the formal model of inflectional rules and lexical entries described in Section 4.

2. VÕRO VERBS SUFFIXES IN CONTEXT

Before presenting a full analysis, the relationship of the Võro verbal suffixes to both the rest of the paradigm and the rest of the word form should be clarified. First, the forms given in Table 3 represent just a fraction of the complete paradigm. Much of the remainder of the suffix paradigm also displays allomorphy. However, this is easily derivable from the forms in Table 3, so these can be understood as reference forms, which stand in for a set of mutually interpredictable forms (Iva 2007: 121). For example, present 3SG Ø both predicts and is predicted by 3PL -*vAq* (*tege⟺tege-väq* 'do'), while present 3SG -*s* predicts and is predicted by 3PL -*sEq* (*elä-s⟺elä-seq* 'live'). A partial paradigm of the synthetic verb forms of a sample verb is given in Table 6, with the reference forms shaded, and the interpredictability of suffix

---

[5] On the other hand, the distinction between classes IX and X does not form part of Iva's classification.

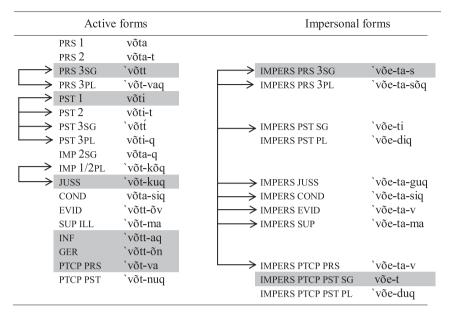| Active forms		Impersonal forms	
PRS 1	võta		
PRS 2	võta-t		
PRS 3SG	ˋvõtt	IMPERS PRS 3SG	ˋvõe-ta-s
PRS 3PL	ˋvõt-vaq	IMPERS PRS 3PL	ˋvõe-ta-sõq
PST 1	võti		
PST 2	võti-t		
PST 3SG	ˋvõtt́	IMPERS PST SG	ˋvõe-ti
PST 3PL	võti-q	IMPERS PST PL	ˋvõe-diq
IMP 2SG	võta-q		
IMP 1/2PL	ˋvõt-kõq		
JUSS	ˋvõt-kuq	IMPERS JUSS	ˋvõe-ta-guq
COND	võta-siq	IMPERS COND	ˋvõe-ta-siq
EVID	ˋvõtt-õv	IMPERS EVID	ˋvõe-ta-v
SUP ILL	ˋvõt-ma	IMPERS SUP	ˋvõe-ta-ma
INF	ˋvõtt-aq		
GER	ˋvõtt-õn		
PTCP PRS	ˋvõt-va	IMPERS PTCP PRS	ˋvõe-ta-v
PTCP PST	ˋvõt-nuq	IMPERS PTCP PST SG	võe-t
		IMPERS PTCP PST PL	ˋvõe-duq

Table 6
Implicative relations between Võro synthetic verb forms ('take'; Iva 2007: 79–80); reference forms discussed in this paper are shaded. (Distinct 1st and 2nd person plural forms are also found in the present and past, formed regularly through suffixation of *-iq* to the corresponding number-neutral form. These may be omitted in the presence of an overt subject pronoun (Iva 2007: 84).)

allomorphs shown with double arrows.[6] Suffixes with no arrows pointing to them display no allomorphy. The impersonal (traditionally 'passive') forms are horizontally aligned with the corresponding active forms.

Second, though the analysis here relies on a distinction between stem and suffix, their boundaries are not always transparent. For clarity of exposition I have made some assumptions here, but have avoided making points of the analysis depend on these assumptions. Following Iva (2007), infinitive *-Aq* and *-En* are assumed to lose their initial vowel when following vowel-final stems; thus consonant-final infinitive *ehitell-äq* 'build' versus vowel-final ˋ*hirnu-q* 'neigh'. The 1st person past suffix *V* is a somewhat more complicated problem, as it corresponds to what are at least superficially different forms. Sometimes it appears to be additive (compare present 3SG ˋ*jää-s* 'stay' with the 1st person past ˋ*jäi*, which has an alternative form ˋ*jäie*), in other cases null (compare present 3SG ˋ*nälgü-s* with the 1st person past ˋ*nälgü* 'starve'), in still others it involves an alternation of the stem-final vowel

---

[6] Not shown in Table 6 are the various other case-marked supine forms beyond the illative (which is the standard citation form of a verb). These can be trivially derived from the illative form, and display no allomorphy.

9

Stem-final vowel	Past tense vowel	
A	i	
E	i	
i	E	
o	o	
VV	Vj	
U	i	*gradation class 'weakening type a' (see Section 3.2)*
U	U	*elsewhere*

Note: /Vj/ is orthographically Vi.

Table 7
Relationship of past tense V to the stem-final vowel found elsewhere (e.g. the 1st person present).

(e.g. present 3SG *istu-s* versus the 1st person past *isti* 'sit'). But in all these cases the form of the past tense suffix is predictable from the stem vowel found elsewhere in the paradigm, as outlined in Table 7. (There are some isolated lexical exceptions.) For convenience we can think of *V* as an abstract suffix which induces a set of morphophonological alternations, without assuming any fixed segmentation of the resulting word form.

The realization of the suffix identified here as *-si* is likewise not always straightforward: with some stems it is geminated (e.g. `*sõim-ssi* 'revile'), while with stems ending in *-s* the segmentation is potentially ambiguous (e.g. the 1st person present ∼ past `*puhksa* ∼ `*puhk-si* (`*puhks-i*?) 'blow'). Following Iva (2007: 121) I make the simplifying assumption that these are all in some sense instances of the same suffix *-si*.

3. SUFFIX ALLOMORPH ASSIGNMENT

The character of Võro verb suffix classes is due to the numerous factors that condition allomorph assignment. These are described in detail below, and can be broadly classified under the headings stem phonology (Section 3.1), stem gradation (Section 3.2) and lexical specification (Section 3.3). This division of labour draws heavily on Blevins's (2007) treatment of conjugation classes in Estonian, although the systems are quite different (in particular, Võro is considerably more complex). One factor that does not seem to play a role is semantic or syntactic properties of the verb, e.g. transitivity or affectedness; the verb classes are purely form-based.

3.1 *Stem phonology*

Much of the allomorphy in Table 3 can be predicted simply on the basis of phonological properties of the stem-final syllable. For example, the

impersonal participle past singular suffix is -*d* with stems ending in -*l* (*tapõl-d* 'kill') or a long vowel (ˋ*jää-d* 'stay') and -*t* elsewhere, e.g. following a short vowel (*kasu-t* 'grow') or diphthong (ˋ*vao-t* 'sink'). Although such predictions can be read directly off a phonological representation of the stem, a certain degree of morphological knowledge is implicit, for two reasons. First, stems may undergo alternations which affect the composition of the stem-final syllable. Second, where we have multiple stems, it may be that the stem which is the best predictor of some allomorph is not the same as the one it attaches to. The crucial point about the predictions described below is not that they describe phonologically MOTIVATED allomorphy (though at times they might), but that they describe phonologically PREDICTABLE allomorphy (Carstairs 1988 and Anderson 2008). That is, given a phonological representation of the stem, no additional lexical specification of the suffix allomorphy is required.

Let us first consider the properties of the stem-final segment. Even a fairly broad phonological characterization is sufficient to account for a good portion of the data. The relevant distinctions are whether the stem-final segment is a (i) short vowel, (ii) long vowel, (iii) diphthong, (iv) *s*, (v) *l* or (vi) a consonant other that *s* or *l*. For stems ending in a short vowel, it will sometimes be necessary to distinguish whether or not the syllable is stressed, which itself may require evaluation of the prosodic foot that it occupies. The alternation patterns in relation to the reference forms are given in Tables 8 and 9. As discussed above (Section 2), the 1st person past is not considered, as the stem–affix division here is ambiguous.

Some predictions can be read directly off the stem that the suffixes attach to: (i) In the present 3SG, a consonant-final stem requires a zero suffix, e.g. ˋ*hirn* 'neigh'. Diphthong-final stems take *s* (ˋ*vao-s* 'sink'). With other stem types, the choice remains open. (ii) The jussive suffix -*guq* is found with stems ending in a long vowel (*jää-guq* 'stay'), diphthong (ˋ*vao-guq* 'sink') or *l* (*tapõl-guq* 'fight'), while -*kuq* is found if the stem ends in a consonant other than *l* (e.g. ˋ*hirn-kuq* 'neigh'). With stems ending in a short vowel either may occur. (iii) The impersonal participle past singular, as mentioned above, is -*d* with stems ending in -*l* or a long vowel (*jää-d* 'stay') and *t* elsewhere. (iv) The present participle suffix -*v* follows a stem ending in a short vowel (*manitsõ-v* 'name'), otherwise it is -*vA* (e.g. consonant-final ˋ*lõp-va* 'end', long vowel-final ˋ*jää-vä* 'stay', and diphthong-final ˋ*vao-va* 'sink').[7]

Predictions for the impersonal present 3SG suffixes could, on the face of it, likewise be read off the stem they attach to, but there is a danger of circularity here, as one of the stem alternations involved (-*As*) seems instead to be

---

[7] As a result, -*vA* and -*v* are in complementary distribution according to stress, with -*vA* appearing after stressed syllables and -*v* after unstressed syllables. But the pattern cannot be attributed directly to stress, as it would predict the possibility of a form such as **mánitsõ̋-va*, where the addition of the syllabic suffix -*vA* to a dactylic stem (*mánitsõ-*) induces secondary stress on the stem–final syllable.

	I*	Ia	2	3	4	5	6	7
PRS 3SG	V	V_1V_2	sk	C	V	V	V	VV
PST I	–	–	–	–	–	–	–	–
INF	V	V_1iV_2	V	V	V	s/l	ts	V + j/vv
GER	V	V_1iV_2	V	V	V	s/l	ts	V + j/vv
JUSS	V	V_1V_2	sk	C	C	s/l	ts	VV
IMPERS PRS 3SG	V	V_1V_2	s	V	V	s/l	V	V + j/vv
IMPERS PTCP PST SG	V	V_1V_2	s	V	V	s/l	V	VV
PTCP PRS	V	V_1V_2	sk	C	C	V	V	VV

V = short vowel, VV = long vowel, C = consonant other than *s* or *l*
*Including diphthongs.

Table 8
Alternations affecting stem-final segments.

due to the suffix allomorph: stems normally ending in a long vowel *VV* are consistently shortened to *Vj* (for front vowels) or *Vvv* (for back vowels) before vowel-initial suffixes, as in the forms of 'stay' shown in Table 9.[8] This circularity can be avoided by basing the prediction on a stem from elsewhere in the paradigm (see e.g. Cameron-Faulkner & Carstairs-McCarthy 2000), namely one that takes a consonant-initial suffix, such as the impersonal participle past singular. This yields the following predictions. The suffix -*As* is found with stems ending in a long vowel (impersonal participle past singular `jää-d⇒impersonal present 3SG jäi-äs 'stay'). The suffix -*dAs* is used for stems ending in *l* (tapõl-d⇒tapõl-das 'fight') and *tAs* for stems ending in a diphthong (`vao-t⇒`vao-tas 'sink') and -*s* (`las-t⇒las-tas 'let'), though note that there is no phonological contrast between lenis *d* and fortis *t* following *s*. For stems ending in a short vowel, what is decisive is whether the suffix follows a (secondarily) stressed syllable: if stressed, -*dAs* is found (érätǘ-däs 'fall behind'), if unstressed, -*tAs* is found (sǘnnü-täs 'be born').[9]

Other predictions likewise require reference to stems other than the one which serves as the base for suffixation. (i) From among the reference forms, the jussive stem turns out to be a better predictor of the infinitive than the infinitive stem itself. Thus, while the infinitive forms `hirnu-q 'neigh', kassu-q 'use' and sõima-duq 'revile' all end in an unstressed short vowel, their jussive stems differ: `hirn-kuq ends in a consonant, kassu-q and sõima-duq each end

---

[8] Iva (2007) treats the /vv/ and /jj/ as part of the suffix, increasing the number of suffix allomorphs. But the lexemes that these consonants occur with form a morphological and phonological class in a number of other respects, so it is more parsimonious to ascribe them to the lexeme itself.

[9] Iva (2007) does not indicate stress in his description. He has, however, confirmed the location of the stresses given here (p.c., October 2011).

13

	1 'grow'	2 'let'	3 'neigh'	4 'end'	5 'fight'	6 'name'	7 'stay'
3SG PRS	kasu	`lask	`hirn	lõpõ-s	`taplõ-s	manitsõ-s	`jää-s
INF	kassu-q	`lasku-q	`hirnu-q	lõppõ-q	tapõl-daq	manits-aq	jäi-äq
GER	kassu-n	`lasku-n	`hirnu-n	lõppõ-n	tapõl-dõn	manits-õn	jäi-en
JUSS	kasu-guq	`lask-kuq	`hirn-kuq	`lõp-kuq	tapõl-guq	manits-kuq	jää-guq
IMPERS PRS 3SG	kasu-tas	las-tas	hirnu-tas	lõpõ-tas	tapõl-das	manitsõ-das	jäi-äs
IMPERS PTCP PST SG	kasu-t	las-t	hirnu-t	lõpõ-t	tapõl-d	manitsõ-t	`jää-d
PTCP PRS	kassu-v	`lask-va	`hirn-va	`lõp-va	`taplõ-v	manitsõ-v	`jää-vä

Table 9
Examples of the patterns in Table 8.

in a vowel. This allows the following prediction: where the jussive stem ends in a consonant – other than *s* or *l* – the infinitive is always *-Aq*; where the jussive stem ends in a vowel, no prediction is made. The suffix *-tAq* is found only with stems in *s* (jussive *helis-kuq*⇒infinitive *helis-taq* 'call'), though there are a few isolated exceptions, given in the following paragraph. With stems in *l*, stress plays a role: if the stem-final syllable is stressed, *-Aq* is found (*éhitél-guq*⇒*éhitéll-äq* 'build'), if unstressed, *-dAq* is found (*tápõl-guq*⇒*tápõl-daq* 'fight'). With stems ending in a long vowel, *-Aq* is found (*jää-guq*⇒*jäi-äq*), while for stems ending in a diphthong, no prediction is possible. (ii) The behaviour of the gerund suffixes *-dEn/-tEn/-En* precisely parallels that of infinitive *-dAq/-tAq/-Aq*. (iii) The segmentation of the 1st person past is ambiguous, but if we again take the jussive stem, two predictions emerge: if this ends in *l* or a long vowel, only *V* is found (jussive *tapõl-guq*⇒1st person past `tapli*; *jää-guq*⇒`jäi). All the phonology-based implications are summarized in Table 10.

There are three verbs that fall outside these generalizations. The verbs *tegemä* 'do' and *nägemä* 'see' have *t*-initial infinitive and gerund forms, which otherwise are exclusively found with s-final stems e.g. infinitive *te-täq*, gerund *te-ten*. The verbs *tulõma* 'come' and *olõma* 'be' have *t*-initial impersonal forms, in spite of its having an l-final stem, e.g. `tul-t, tul-tas*. As these are high-frequency core lexemes, I will assume that these forms are exceptionally specified in the lexical entry (see Section 4.4 below).[10]

Factoring in the generalizations in Table 10 allows for a greatly simplified picture of the inflectional classes, with only seven remaining out of the original 23 as given in Table 3. Adapting what is already the established tradition in descriptions of Võro (e.g. Keem 1997), the primary distinction is due to the cross-classification of the suffix allomorphs for present 3SG (here 'A' vs. 'B'), which then cross-classifies with the past tense allomorphs (here '1' vs. '2'), which yields the four macroclasses A1, A2, B2, B1.[11] In the case of A1 and A2 nothing more need be said, as the rest of the paradigm is now predictable. For B1, the subtypes are due to allomorphy of the infinitive and gerund, though note that since the infinitive and gerund here are mutually interpredictable, it is in effect a single choice. The subtypes of B2 are due to allomorphy of the gerund and jussive. In the next section, we see that much of the allomorphy in Table 11 can in turn be derived from other aspects of stem alternation, namely the gradation system.

---

[10] There is no frequency data available for Võro, but if we take the data from standard written Estonian as a rough guide, the corresponding verbs are ranked as the first (*olema*), third (*tulema*), fifth (*tegema*) and eleventh (*nägema*) most frequent verb lexemes (Kaalep & Muischnek 2002).

[11] One reason for taking the present 3SG as primary is that the *-si* suffix has lately shown a tendency to spread beyond its original bounds (Iva 2007: 88), likely due to influence from Standard Estonian, where it is the default past tense suffix.

15

	C-final	s-final	l-final		V-final		VV-final	diphth-final
			+ stress	− stress	+ stress	− stress		
PRS 3SG	Ø	*	*	*	–	–	–	s
PST I	–	–	V	V	–	–	V	–
INF	Aq	tAq	Aq	dAq	–	–	Aq	–
GER	En	tEn	En	dEn	–	–	En	–
⇨ JUSS	kuq	kuq	guq	guq	–	–	guq	guq
IMPERS PRS 3SG	*	tAs	dAs	dAs	dAs	tAs	As	tAs
⇨ IMPERS PTCP PST SG	*	t	d	d	t	t	d	t
PTCP PRS	vA	*	*	*	v	v	vA	vA

* = does not occur in these contexts (see Table 8)

Table 10
Suffix assignment predictable from phonology of the stem-final syllable; PST I, INF, GER, JUSS predictions based on JUSS stem (indicated by the upper shaded block), IMPERS PRS 3SG, IMPER PTCP PST SG predictions based on IMPER PTCP PST SG stem (indicated by the lower shaded block).

	A1	A2	B1		B2		
			a	b	a	b	c
PRS 3SG	Ø	Ø	s	s	s	s	s
PST 1	V	si	V	V	si	si	si
INF	Aq	Aq	Aq	dAq	dAq	dAq	dAq
GER	En	En	En	dEn	dEn	tEn	tEn
JUSS	guq	guq	guq	guq	guq	guq	kuq

Table 11
Classes that result after the predictions in Table 10 are factored out. The cells that distinguish the
subclasses of B1 and B2 are shaded.

3.2 Stem gradation

In common with other Balto-Finnic languages, Võro exhibits a system
of inflectional stem alternations in both the verbal and nominal systems
known as gradation, in which a so-called strong grade stem alternates
with a weak grade stem. The phonological manifestations of the
opposition are quite varied, involving both prosodic and segmental alter-
nations. A key role is played by the three-way quantity opposition, which
Võro shares with standard Estonian. This warrants at least brief exposition
here.

Stressed syllables may be short (quantity 1, abbreviated 'Q1'), long
(quantity 2, abbreviated 'Q2'), or overlong (quantity 3, abbreviated 'Q3').
Q1 syllables are necessarily open, while Q2 and Q3 syllables may be closed or
open. These in turn serve as the heads of prosodic feet: a Q1 or Q2 foot
necessarily consists of a stressed syllable followed by one or two unstressed
syllables, while a Q3 syllable can constitute a foot on its own – and, by im-
plication, all monosyllabic words are necessarily Q3, in order to constitute a
complete prosodic foot. Primary stress is typically initial; since prosodic
words must be exhaustively parsed, secondary stress may also occur. Table 12
gives some examples of verb stems characterized by their (final) prosodic
foot. Note that the fortis consonants *k*, *p* and *t* are construed phonologically
as (short) geminates which close a preceding syllable; thus *siba-* is a Q1 stem
while *hupa-* is a Q2 stem.

In the strong grade∼weak grade alternation, the weak grade is either
shorter in terms of syllable quantity, and/or undergoes lenition, deletion or
assimilation of the consonant at the boundary between the first and second
syllables, sometimes with further changes to the stem vowel or syllable
structure. Some examples are shown in Table 13. The stems in (a–d) all have
a quantity alternation; the stems in (d–g) display consonant assimilation or
deletion (*nd∼nn*, *k∼Ø*, *g∼Ø*). (The consonant alternations in (b–c) are

16

Q1 stem		Q2 stem		Q3 stem	
siba-guq	'scurry'	hupa-guq	'lift'	`uppu-guq	'drown'
kosi-guq	'grow'	hooba-guq	'row'	`laabu-guq	'prosper'
elä-guq	'live'	palga-kuq	'employ'	`nälgü-guq	'starve'

Table 12
Three degrees of foot quantity, illustrated with jussive forms.

	Strong		Weak		
a. 'curse'	`häägä	(PRS I)	häägä-daq	(INF)	*Q3 ∼ Q2*
b. 'push'	`tuuka	(PRS I)	tuuga-daq	(INF)	*Q3 ∼ Q2*
c. 'scurry'	sipa	(PRS I)	siba-ma	(SUP ILL)	*Q2 ∼ Q1*
d. 'price'	`hinda	(PRS I)	hinna-daq	(INF)	*Q3 ∼ Q2, consonant assimilation*
e. 'blow'	`puhksa	(PRS I)	puhas-taq	(INF)	*consonant deletion*
f. 'shear'	pügä-mä	(SUP ILL)	`püä	(PRS I)	*consonant deletion*
g. 'wheeze'	kryyga	(PRS I)	krõõa-daq	(INF)	*consonant deletion*

Table 13
Some examples of stem gradation.

concomitant with the quantity alternation, while the quantity alternation in (e–f) is a consequence of the syllable structure changes.)

As the examples show, stem grade is a morphological, not a phonological property. Phonologically comparable stem types may function as strong or weak grade, depending on the verb. For example, both *siba-* and *pügä-* are Q1 stems, but *siba-* functions as a weak stem and *pügä-* as a strong stem. Similarly, *tuuga-* and *kryyga-* are both Q2 stems with a long vowel and a lenis consonant at the syllable boundary, but *tuuga-* is a weak stem and *kryyga-* a strong stem. But since strong and weak grades are relative properties, given any pair of stem alternants, it is always possible to identify which is strong and which is weak.

Some stem grade alternations are exclusively grammatically conditioned, applying across the whole lexicon, while others are restricted to particular groups of lexemes. It is these lexically conditioned patterns of stem alternation which can, to a large extent, predict the residual suffix allomorphy unaccounted for by the phonological implications outlined in Section 3.1. Table 14 identifies the major stem grade alternation classes, in terms of the present 3SG, the past 1st person and the jussive. These are sufficient to identify the stem gradation pattern. The full patterns in terms of the reference forms

17

	Weakening		Strengthening	Geminating		Invariant	
	Type a 'shear'	Type b 'be born'	'revile'	Type a 'tire'	Type b 'scurry'	QI stem 'grow'	Q3 stem 'starve'
PRS 3SG	**pügä**	sünnü-s	**`sõima-s**	väsü-s	**sipa**	kosi	`nälgü-s
PST I	`pöi	**`sündü**	**`sõim-ssi**	**`vässü**	siba-si	kosõ	`nälgü
JUSS	**pügä-guq**	**`sündü-guq**	sõima-kuq	väsü-guq	siba-guq	kosi-guq	`nälgü-guq

Table 14
Stem gradation classes; strong grade stems shown in boldface.

are given in Appendix A. The names given to the classes in Table 14 are adapted from the terms traditional in Balto-Finnic studies.[12]

There are five classes which display gradation, plus a sixth invariant class which lacks it.[13] The weakening class is distinguished by having a strong grade jussive. Within it there are two subtypes: type a has a strong present 3SG and a weak past 1st person, while type b has the reverse pattern. All the other classes (except the invariant class) have a weak jussive, and are distinguished by the behaviour of the present 3SG and the past 1st person. In the strengthening class both of these are strong. Geminating type a has weak grade in the present 3SG and strong grade in the past 1st person, while type b has the reverse pattern. The distribution of lexemes across these classes is to some extent related to their prosodic properties, in particular, the distribution of Q1 stems. These are largely absent in the weakening and strengthening classes (there are a few among the weakening type a class), while the geminated classes consist exclusively of verbs whose weak stem is Q1. The invariant type includes Q1 and Q3 stems, as well as Q2 stems with a transparent word-forming suffix (e.g. *hirnahta-s* 'neigh.3SG.PRS', whose stem ends in the word-forming suffix -*ta*).

The five classes that display a gradation alternation each occur almost exclusively with a single pattern of suffixation. Therefore, if the gradation alternation of a verb is known, its suffixation pattern is in most cases completely determined. Table 15 lists which suffixes are predicted by which gradation class, and identifies which of the suffix classes in Table 11 above the resulting pattern corresponds to. With three of the gradation classes the prediction is absolute, with the two of them (the strengthening class and the geminating type a) it is nearly absolute, leaving one pair of competing allomorphs unresolved in each case.

As with the phonology-based predictions described in Table 10, these are not necessarily local predictions that obtain between a given stem alternant and the suffixes that attach to it. For example, the present 3SG has a strong stem in both the weakening type a class and the strengthening type b class, but in the former class the suffix is predicted to be Ø, and in the latter it is predicted to be -*s*. Rather, the predictions are based on the gradation paradigm as a whole, and not on individual forms.[14]

---

[12] The terminology is derived from the behaviour of noun stems, referring to the gradation of the genitive stem compared to the nominative. Roughly speaking, 'weakening' stems were originally vowel–final and 'strengthening' stems consonant–final.

[13] As noted in Appendix A, such verbs may in fact undergo gradation alternations, but these are determined by general morphosyntactic and phonological properties, and not lexically specified.

[14] To some extent predictions in the other direction are possible, from suffixation to stem gradation, but the coverage is worse. Thus, the stem gradation⇒suffix class predictions in Table 15 account for about 30% of the lexicon (1393/4642), provided we allow for the indeterminacy in the jussive of the strengthening class and in the 3SG of the geminating type

	Weakening		Strengthening	Geminating		Invariant
	Type a	Type b		Type a	Type b	
PRS 3SG	Ø	s	s	–	Ø	–
PST I	V	V	si	V	si	–
INF	Aq	dAq	dAq	Aq	Aq	–
GER	En	dEn	tEn	En	En	–
JUSS	guq	guq	–	guq	guq	guq
IMPERS PRS 3SG						
IMPERS PTCP						
PST SG						
PTCP PRS						
Corresponds to suffix classes:	A1	B1b	B2c, B2b	B1a, A1	A2	A1, B1b, B1a, B2a

Table 15
Stem gradation ⇒ suffix implications. Cells which are fully determined by stem phonology (Table 10) are shaded.

3.3 *Lexical specification*

Factoring in the predictions that can be arrived at by consideration of the phonology of the stem-final syllable, and the pattern of stem gradation, we arrive at the highly restricted inventory of suffix allomorph choices in Table 16, which represents the contexts still open to lexical specification. While the predictions made so far have been categorical, at this point we can consider the role of statistics in order to determine what is the default pattern, and what must be lexically specified. For each value there is a choice of two allomorphs, and in most cases the distribution across the lexicon is highly imbalanced. The type count for the suffixes in Table 16 is given in Table 17, based on the word list provided by Iva (2007). Note that this does not count the lexicon as a whole, but only those contexts in which the suffix allomorph has not been determined by stem phonology or gradation pattern.

In the case of the present 3SG and past 1st person, -s and V are the clear favourites, as indeed they are across the lexicon as a whole. The jussive suffix

---

a class. In the other direction, only about a third of the lexicon is covered (719/4642), as follows:

A2 ⇒ geminating b	88	lexemes
B2a ⇒ invariant	185	lexemes
B2b ⇒ strengthening	9	lexemes
B2c ⇒ strengthening	437	lexemes

And note that this predicts only the pattern of gradation, not how gradation is realized phonologically. As pointed out in Table 13, this may vary.

20

	Weakening		Strengthening	Geminating		Invariant
	Type a	Type b		Type a	Type b	
PRS 3SG				s, Ø		s, Ø
PST I						V, si
INF						dAq, Aq
GER						dEn, En
JUSS			kuq, guq			
IMPERS PRS 3SG						
IMPERS PTCP PST SG						
PTCP PRS						
Corresponds to suffix classes:			B2c, B2b	B1a, A1		A1, B1b, B1a, B2a

Table 16
Contexts in which lexical specification of suffixes is possible. Cells which are fully determined by stem phonology (Table 10) or stem gradation (Table 15) are shaded.

-*kuq*, though exceptional in the lexicon as a whole (it is only found after C-final or *s*-final stems), is clearly the default for the strengthening class, the one context where it can be lexically specified. In all these cases there is one allomorph that accounts for 93% or more of the type frequency. If we assume that this one is simply assigned by default, only the exceptional suffix allomorphs need be encoded in the lexical entry.

For the infinitive and gerund the picture is less clear, as the *A*-initial allomorphs and the *d*- initial allomorphs are more or less equally distributed. This can be resolved if we consider not just frequency, but ask which one applies in the most heterogeneous set of contexts. If we look more closely at the phonological properties of the stem, there is again a clear imbalance in the distribution of the allomorphs. 92% (875/954) of the verbs that take -*Aq*/-*En* have stems terminating in -*tA* (or -*dA* when following a resonant). Looked at from the other side, all verb stems in the lexicon that end in -*tA*/-*dA* take -*Aq*/-*En*, so there is a strong generalization here. In one sense this is a morphological or lexical fact, since this stem termination is in many cases identifiable as a productive word-forming suffix, found with both denominal/deadjectival verbs (*külme-tä*- 'freeze' < *külm* 'cold') and deverbal verbs (*sünnü-tä*- 'give birth' < `*sündü*- 'be born'). But the phonological generalization captures the implication without necessarily having to recognize any stem-internal morphological constituents. The phonological implications from Table 10 can thus be enriched to consider the quality both of the stem-final vowel and of the preceding segment: if the (jussive) stem ends in a dental stop (*t* or *d*) plus *A*, then the infinitive is -*Aq* and the gerund is -*En*.

21

PRS 3SG	s (3032)	versus	Ø (89)	97% > 3%
PST I	V (2618)	versus	si (185)	93% > 7%
INF/GER	Aq/En (954)	versus	dAq/dEn (771)	55% ≈ 45%
JUSS	kuq (437)	versus	guq (9)	98% > 2%

Table 17
Type frequency of suffixes in Table 16, showing number of lexemes with a given suffix.

If we factor out these items from the total, then *-Aq/En* occurs with 79 lexemes (9%) and *-dAq* with 771 (91%), making *-dAq* the clear default choice.

Sensitivity to further instances of phonologically transparent word-forming morphology predicts the assignment of still more instances of non-default suffixes. The stem-final sequences *-nE-* and *-hA-* also correspond to word-forming suffixes.[15] These have the property of appearing only in certain portions of the paradigm, as shown in Table 18. Because of this, and because of some phonological overlap with other stem types, the one consistent predictor from among the reference forms is the present participle.[16] With this as the point of departure, we can then say that verbs whose present participle stem ends in *-hA-* or *-nE-* (*harinõ-v*, *sibaha-v*) always take the past 1st person suffix *-si*. This predicts ALL the remaining instances of *-si*, which thus need not be lexically specified at all. Further, stem-final *-nE-* consistently predicts jussive *-guq*. And the small set of verbs of the strengthening gradation class that exceptionally take jussive *-guq* in place of *-kuq* all have the *-nE-* suffix (with one single exception), so these are accounted for as well.

If we factor in these enriched phonological predictions, the possibilities for lexical specification of suffixes are highly restricted. At most there are three possible types, differing in terms of which cells remain open for lexical specification: (i) verbs which lack stem gradation,[17] for which the present 3SG, the infinitive and the gerund can be lexically specified (Table 19, a); (ii) long-vowel–final stems and verbs of the geminating type b gradation class, for which the PRESENT 3SG can be lexically specified (Table 19, b); and (iii) diphthong-final stems, for which the infinitive and gerund can be lexically specified (Table 19, c). Underlyingly, however, all three types share a common trait: either all the lexically specified suffix allomorphs are exceptional, or none are. Therefore, we need recognize just one class of lexemes that are specified as having exceptional allomorphs.

---

[15] Verbs with the *-hA* suffix constitute an alternative inflectional paradigm for verbs of the geminating type a gradation class. The two types also differ in their affixation, independent of the distribution of *-hA-*; compare the *-hA-* verb forms *sibaha-s* (PRS 3SG) and *siba-daq* (INF) with the corresponding forms of the geminating type a forms: *sipa*, *sipa-q*.

[16] Some verb stems end in *-ha* in the present tense; these belong to the weakening class and have a stem gradation pattern with a strong stem ending in *-hC* and a weak stem in *-h*, e.g. infinitive `*tahta-q* ∼ present 3SG *taha* 'want'.

[17] On the behaviour of the infinitive and gerund stem of the *kosima* type, see Appendix A.

	'get used to'	'scurry'	Associated predictions
PRS 3SG	harinõ-s	sibaha-s	
PST I	hari-si	siba-si	Stem-final -*nE-* ⇒ past I -*si*
INF	hari(nõ)-daq	siba-daq	jussive -*guq*
GER	hari(nõ)-dõn	siba-den	
JUSS	hari-guq	siba-guq	Stem-final -*hA-* ⇒ past I -*si*
IMPERS PRS 3SG	hari(nõ)-das	siba-tas	
IMPERS PTCP PST SG	hari(nõ)-t	siba-t	
PTCP PRS	harinõ-v	sibaha-v	

Table 18
Alternating word formation suffixes.

By this reckoning, at most 2% of the verb lexemes in the lexicon (88/4642) need to be lexically specified as taking exceptional suffix allomorphs.[18] Consider this in the light of the question posed towards the beginning of this paper: What does knowledge of a rampantly crossclassifying system such as the 23 suffix classes of Võro actually constitute? Largely it constitutes knowledge of stem phonology, and of the system of stem alternations known as gradation, and the mapping relations between these and the suffix allomorphs. Explicit knowledge of suffix distribution – that is, what needs to be known about suffix allomorphy independent of the rest of the morphological system – can be reduced to this minimal stipulation.

4. FORMAL REPRESENTATION

The description of suffix allomorphy outlined in Section 3 mixes diverse elements, ranging from the phonologically transparent predictions, through morphophonology of varying degrees of opacity, all the way to direct lexical specification. On a strictly modular view of language, a formal analysis of this system ought to handle these elements separately – at the very least with distinct phonological and morphological rules. But these elements are so interwoven that such a strict segregation would give an entirely false picture. The analysis offered here aims to represent the mixed properties of the system, using composite rules that make reference to the various phonological and morphological conditions.

The formal representation is constructed Network Morphology, using the computational language DATR (Corbett & Fraser 1993, Evans &

---

[18] Alongside the four completely irregular verbs discussed in Section 3.1.

23

a. 3SG PRS/INF/GER		b. 3SG PRS		c. INF/GER	
Default 'starve'	Exceptional 'court'	Default 'stay'	Exceptional 'become'	Default 'sink'	Exceptional 'sink'
PRS 3SG	`nälgü-s	kosi	`jää-s	`saa	`vao-s	`vao-s
PST I	`nälgü	kosõ	`jäi	`sai	`vaio	`vaio
INF	`nälgü-daq	kossi-q	jäi-aq	sai-aq	`vao-daq	vaio-q
GER	`nälgü-den	kossi-n	jäi-en	sai-õn	`vao-den	vaio-n

Table 19
Types of lexical specification of suffixes; cells for which the allomorph is determined by other factors are shaded.

Gazdar 1996, Brown & Hippisley 2012). Similar to Paradigm Function Morphology, this is a default inheritance model, in which inflectional rules are arranged in a hierarchy of specificity from general categories (such as word classes) down to individual lexical entries. Its suitability for the present task is due to its simplicity, and to the fact that it has already been successfully used to represent inflectional class systems of the 'well-behaved' inflectional class systems (those which come close to the ideal of the Paradigm Economy Principal), making for an instructive comparison with the Võro system.

By way of a brief introduction, let us see how the formalization of a simple inflectional class system would look, taking the small fragment of Latin illustrated above in Table 4, repeated as Table 20. Recall that there are three distinct paradigm types. The *agricola* type is distinct from the other two in nearly all its forms. However, *discipulus* and *puer* differ from each other only in the nominative singular. As this difference is phonologically predictable (stems ending in /r/ take a zero suffix), they can be treated as subtypes of a single inflection class. In the analysis which follows the 1st declension (the *agricola* type) will be treated as the default class. This has no particular linguistic justification, but will help illustrate an important property of the formal machinery.

The model of the inflectional system is made up of five nodes, each housing rules that are structured as attribute/value pairs separated by ' = = '. The attribute, on the left-hand side enclosed in ' < > ', can be understood as a feature, and the value on the right-hand side can be understood as a value of that feature. The model presupposes that lexical entries minimally contain information about inflectional class and about the phonology of the stem-final syllables. The inflectional rules can then be thought of as a query which looks for the values of a feature suffix, taking the lexical entry as its starting point.

24

	Suffixes	1st declension 'wax'	2nd declension Stem in -r 'boy'	Other stems 'disciple'
NOM SG	a, Ø, us	cēr-a	puer	discipul-us
ACC SG	am, um	cēr-am	puer-um	discipul-um
GEN SG	ae, ī	cēr-ae	puer-ī	discipul-ī
DAT SG	ae, ō	cēr-ae	puer-ō	discipul-ō
ABL SG	ā, ō	cēr-ā	puer-ō	discipul-ō

Table 20
Conditioned allomorphy within the Latin 2nd declension. (=Table 4)

The top node NOUN in (1) contains two rules. The first directs the query to the node SUFFIXES, where the value of the inflection class feature `class suffix` is evaluated. The second simply states that the default value of any feature, if not otherwise defined, is null.

(1) NOUN:
```
<suffix> == SUFFIXES:<"<class suffix>">
<> ==.
```

The node SUFFIXES in (2) partitions the rules into two groups: lexemes with the `class suffix` value ii (for 2nd declension) are directed to the node DECLENSION_II. Everything else is directed to DECLENSION_I, reflecting the assumption that this is the default class. That means that in the associated lexicon, only membership in the 2nd declension needs to be overtly encoded; class membership of 1st declension lexemes can be left underspecified.

(2) SUFFIXES:
```
<> == DECLENSION_I:<>
<ii> == DECLENSION_II:<>.
```

The description of suffix forms at DECLENSION_I and DECLENSION_II in (3) is largely straightforward, the real point of interest here is the treatment of the nominative singular of the 2nd declension (ultimately based on Brown & Hippisley 1994).[19] Rather than defining a form, it induces a further evaluative path, directing the query to the node NOM_SG in (4), where the

---

[19] Of course, the syncretism of the dative with the genitive (1st declension) or ablative (2nd declension) would have to be addressed in a fuller analysis. This is not directly relevant to the issue at hand, and so has been left unresolved, as it would take the discussion off on a tangent.

stem-final segment is evaluated. If this is /r/, a null suffix is selected; other-wise, *-us* is selected.

(3) ```
     DECLENSION_I:
         <nom sg> == a
         <acc sg> == am
         <gen sg> == ae
         <dat sg> == ae
         <abl sg> == aa.
     DECLENSION_II:
         <nom sg> == NOM_SG:<"">
         <acc sg> == um
         <gen sg> == ii
         <dat sg> == oo
         <abl sg> == oo.
```

(4)  ```
     NOM_SG:
         <> == us
         <r> == 0.
```

This Latin fragment illustrates two ways in which inflectional allomorphs can be assigned: (i) through lexical specification, expressed here as a value of `class`, or (ii) through evaluation of some other property of the lexeme – in this case, stem phonology. The analysis of Võro conjugation offered below makes use of these same devices (though as we shall see, with some important differences compared to the Latin fragment). The following sections lay out the major properties of the model. For reference, the full model is given in Appendices B and C.

4.1 General properties of the model

An abbreviated view of the central node of the network, VERB, is shown in (5). This is where general properties of the system are housed. The model generates only the final portion of the inflected word (here termed end), namely the stem-final segment plus suffix (line 1 of (5)). The next two lines state phonological properties: by default, the stem-final segment is a vowel (the '$' symbol indicates that this is a variable whose values can be further specified), and the stem-final syllable is unstressed. The final line assigns a null value to any feature in case no other rule applies.

(5) ```
     VERB:
1   <end> == "" "<suffix>"
2   == $vowel
3   <stress stem_final> == unstressed
4   <> ==
         ...
```

26

The output of these rules is an underlying morphological representation, whose translation into surface-true forms may involve additional morphophonological rules which are not explicitly treated here. These include: (i) the quality of past tense -*V* (see Table 7 above) and (ii) the reduction of a final -*i* in the past tense 3SG forms of some stem types to palatalization of the preceding consonant, e.g. *kõnõli* (1st person) ∼ *kõnõl̃* (3SG) 'talk'.

### 4.2 *Three types of allomorph assignment*

In line with the presentation in Sections 3.1–3.3, the sets of rules for the individual reference forms can be divided into three types, of increasing complexity: (i) those which make reference to stem phonology alone, (ii) those which make reference to stem phonology and stem alternation type, and (iii) those which make reference to stem phonology, stem alternation type, and lexical specification. In the following sections each type is described in turn.

#### 4.2.1 *Type 1: Stem phonology alone*

The impersonal participle past singular is a form determined solely by the phonology of the stem-final syllable. Examples (6)–(8) give the sequence of steps.

(6)  VERB:
```
    ...
    <suffix impers ptcp sg> == IMPERS_PTCP:<"<segment
     stem_final impers ptcp>">
    ...
```

(7)  IMPERS_PTCP:
```
    <l> == FORM_DENTAL
    <$long_vowel> == FORM_DENTAL
    <> == FORM_DENTAL:<fortis>.
```

(8)  FORM_DENTAL:
```
    <null> ==
    <fortis> == t
    <> == d.
```

At the node VERB in (6) the rule says to go to the node IMPERS_PTCP, in (7), and select the appropriate instruction on the basis of the stem-final segment. At this node there are three instructions, all of which refer to a third node FORM_DENTAL, in (8), which provides the actual form: stems ending in /l/ or a long vowel take lenis /d/; otherwise the fortis /t/ is selected.

#### 4.2.2 *Type 2: Stem phonology plus stem gradation*

The past 1st person is a form that requires reference both to the phonology of the stem-final syllable and to the stem gradation pattern, which obviously leads to a more complex sequence of rules, laid out in (9)–(12).

(9)   VERB:
```
      ...
      <suffix pst 1> == PST_NODE1:<"<segment stem_final
       juss>">
      ...
```

(10)  PST_NODE1:
```
      <l> == FORM_PST
      <$diphthong> == FORM_PST
      <$long_vowel> == FORM_PST
      <> == FORM_PST:<"<class stem>">
      <a> == PST_NODE2:<"<phon segment stem_penult ptcp
       prs>">
      <e> == PST_NODE3:<"<phon segment stem_penult ptcp
       prs>">.
```

(11)  PST_NODE2:
```
      <h> == FORM_PST:<strengthening>
      <> == FORM_PST:<"<class stem>">.
    PST_NODE3:
      <n> == FORM_PST:<strengthening>
      <> == FORM_PST:<"<class stem>">.
```

(12)  FORM_PST:
```
      <strengthening> == s i
      <geminating b> == <strengthening>
      <> == vowel.
```

At the VERB node in (9), the query is directed to PST_NODE1 in (10), where the stem-final vowel of the jussive is evaluated. For stems ending in /l/, a diphthong or a long vowel, no further evaluation is performed: the default form -*V* at the node FORM_PST in (12) is assigned. For other stem-final segments – except for /a/ and /e/ – the form is determined by the stem gradation pattern at FORM_PST. The stem gradation pattern is identified here through the CLASS STEM feature whose values are strengthening, weakening a, weakening b, geminating a and geminating b. At FORM_PST the suffix -*si* is assigned if the value is strengthening or geminate b, the suffix -*V* elsewhere.

Stems ending in /a/ or /e/ require an additional level of evaluation. Recall from Section 3.3 above that the words formed with the suffixes -*ha* and -*ne* invariably take the past tense suffix -*si*, regardless of their stem gradation

28

patterns. This is modelled through two additional nodes, `PST_NODE2` and `PST_NODE3` in (11), which evaluate the stem-penultimate consonant just in case the stem ends in -*a* or -*e*. The suffix -*si* is consequently assigned to stems in -*ha* and -*ne*, while suffix assignment on the basis of the stem gradation pattern is allowed elsewhere.

### 4.2.3 *Type 3: Stem phonology plus stem alternation plus lexical specification*

The infinitive is a form whose assignment depends on all three components: stem phonology, stem gradation pattern and lexical specification. The relevant rules are presented in (13)–(18).

(13) `VERB:`

    ...

    `<suffix inf> == INF_NODE1:<"<segment stem_final juss>"> a q`

    ...

(14) `INF_NODE1:`

    `<$long_vowel> == FORM_DENTAL:<null>`

    `<$consonant> == FORM_DENTAL:<null>`

    `<$diphthong> == FORM_DENTAL:<"<class suffix>">`

    `<l> == INF_NODE2:<"<phon stress stem_final juss>">`

    `<s> == FORM_DENTAL:<fortis>`

    `<a> == INF_NODE3:<"<phon segment stem_penult ptcp prs>">`

    `<> == INF_NODE4:<"<class stem>">.`

(15) `INF_NODE2:`

    `<stressed> == FORM_DENTAL:<null>`

    `<> == FORM_DENTAL.`

(16) `INF_NODE3:`

    `<t> == FORM_DENTAL:<null>`

    `<> == INF_NODE4:<"<class stem>">.`

(17) `INF_NODE4:`

    `<weakening a> == FORM_DENTAL:<null>`

    `<weakening b> == FORM_DENTAL`

    `<strengthening> == FORM_DENTAL`

    `<geminating> == FORM_DENTAL:<null>`

    `<> == FORM_DENTAL:<"<class suffix>">.`

(18) `FORM_DENTAL:`

    `<null> ==`

    `<fortis> == t`

    `<> == d.`    (=(8) above)

The initial rule at the VERB node in (13) consists of two components. Recall that the allomorphs of the infinitive share the terminal formative /Aq/. This much can then be defined at the VERB node, so that the rules only need account for the initial formative. This formative is assigned at the node FORM_DENTAL, the same one that was used above for the impersonal participle past singular. But because the assignment rules themselves are different, the distribution of the allomorphs does not necessarily coincide, and the two cannot be conflated.

In the first step, the stem-final segment of the jussive is evaluated at INF_NODE1 in (14), where there are five possible outcomes: (i) if the stem ends in a long vowel or consonant other than /l/, a Ø form is assigned at FORM_DENTAL (in (15), repeated here from (8) above); (ii) if the stem ends in /l/, the stress of the stem-final syllable is evaluated at INF_NODE2 in (15)); (iii) if the stem ends in /a/, the stem-penultimate segment is evaluated at INF_NODE3 in (16); (iv) if the stem ends in a diphthong, lexical specification, in the guise of the feature class suffix, is evaluated at FORM_DENTAL; and (v) otherwise, the stem gradation pattern is evaluated at INF_NODE4 in (17).

Most of this rule sequence involves operations already seen above, namely evaluation of the phonological properties of the stem-final syllable and the stem-gradation pattern. But importantly, there is also lexical specification of the suffix, which occurs at two points in the rule sequence (highlighted in (14) and (17)). As pointed out above (Section 3.3), we need distinguish only two inflection classes, one of which is clearly the default, and one of which is exceptional, here given the name null, in recognition of the fact that it is characterized by a Ø exponent.

### 4.3 *Relations between forms*

#### 4.3.1 *Mutual implicature*

So far the analysis of allomorphy has only addressed isolated reference forms. But as pointed out above in Table 6, the reference forms sometimes stand in for a set of mutually interpredictable forms. There are two basic patterns: (i) the interpredictability coincides with morphological identity, and (ii) the interpredictability is independent of morphological identity.

The first pattern falls out from segmenting suffixes into a variable portion, subject to allomorphy, and an invariant portion, as demonstrated above with the infinitive. For example, the past tense forms are mutually implicative because they all begin with the same formative – equivalent to the 1st person suffix described above – followed by a common set of person-number suffixes, e.g. *kosõ ~ kosõ-t ~ kosõ-q* 'court' (1st person/3SG ~ 2nd person ~ 3PL) versus *siba-si ~ siba-si-t ~ siba-si-q*

30

|  | 'live' | 'do' |
|---|---|---|
| PRS 3SG | elä-s | tege |
| PRS 3PL | elä-seq | tege-väq |

*Table 21*
3rd person singular and plural forms.

'scurry'.[20] To account for this we can recast the 1st person past rule in (9) above as a generic past tense rule, as in (19). The suffix is broken up into two formatives, the first realizing the tense value 'past', the second (called `termination` here) realizing the individual person number values.

(19) VERB:
```
      ...
      <suffix pst> == PST_NODE1:<"<segment stem_final
       juss>"><termination pst>
      <termination pst 2> == t
      <termination pst 3 pl> == q   (=(9), revised)
      ...
```

The second pattern is evident in the present 3rd person, shown in Table 21, where the mutual implicature of singular and plural forms cannot readily be reduced to a common formative. Thus, even if we treat the singular forms as a base (as historically it was), the plural forms still involve distinct terminations in the two classes (*-Eq* vs. *-vAq*). This is the one instance of bona fide inflectional class behaviour that we have in the Võro verbal paradigm, in that the implicature between the forms must be stated explicitly.

Modelling this requires a certain elaboration of the rules, as it is a pair of allomorphs that must be selected. Therefore, the rule sequence in (20)–(22) must select not a single allomorph, but some placeholder feature whose value can in turn be used to select the individual allomorphs.

(20) VERB:
```
      ...
1   <index> == PRS_NODE1:<"<segment stem_final prs 3
       sg>">
2   <suffix prs 3> == FORM_PRS:<"<index>">
      ...
```

---

[20] While the past tense forms all share the same suffix formative, they do not necessarily share a stem, as different stem grades may be involved. For example, weakening type b stems have weak grade in the 1st and 2nd person and strong grade in the 3rd; see Table A1 in Appendix A.

31

```
(21) PRS_NODE1:
     <$consonant> == null
     <$diphthong> ==
     <> == PRS_NODE2:<"<class stem>">.
   PRS_NODE2:
     <weakening a> == null
     <geminating b> == null
     <> == "<class suffix>".
```

```
(22) FORM_PRS:
     <> == CLASS_DEFAULT:<>
     <null> == CLASS_NULL:<>.
   CLASS_DEFAULT:
     <sg> == s
     <pl> == s e d.
   CLASS_NULL:
     <sg> ==
     <pl> == v a d.
```

Thus, starting at the VERB node in (20), line 1, the value of this feature (index) is determined through evaluation of the stem-final syllable, stem gradation pattern and lexically-specified class suffix feature in (21). The resulting value then splits the lexicon into two mini-inflection classes, as in (22), rather in the same way as the Latin example above (see (2)–(4) above) – though the two examples contrast in a revealing way. In the Latin example the primary split is lexically specified; phonology-based allomorph assignment (nominative singular -Ø versus -*us*) is restricted to one of the inflection classes. In Võro the hierarchy of rules is reversed: lexical specification is subordinate to phonology-based rules.

In both the above examples – the past tense forms and the present tense 3rd person forms – the domain within which mutual implicature obtains coincides with a morphosyntactic natural class, namely past tense or 3rd person. But this need not necessarily be the case. For example, the jussive and imperative plural are mutually implicative because they share allomorphy of the initial velar (compare the jussive/imperative plural pair *elä-guq* ~ *elä-geq* 'live' with *sõima-kuq* ~ *sõima-kõq* 'revile').[21] But there is no obvious way to reduce these to a single rule, because the two are morphosyntactically unrelated, at least in terms of the feature structure assumed here. Therefore they are treated as two distinct rules at the VERB node in (23), which nonetheless both make reference to the same sequence of allomorph-selecting rules, starting at JUSS_NODE1.

---

[21] The vowel alternation is due to vowel harmony.

(23) `VERB:`

```
    ...
    <suffix juss> == JUSS_NODE1:<"<segment stem_final
     juss>"> u q
    <suffix imp pl> == JUSS_NODE1:<"<segment stem_
     final juss>"> e q
    ...
```

Thus, the initial velar segment in both forms is construed as morphologically identical; but in contrast to the treatment of the past tense, this identity is morphologically stipulated rather than morphsyntactically derived.

4.3.2 *Partial implicature*

The infinitive and gerund have been treated here as distinct reference forms, but in fact the realization of the initial element of the suffix (/d/, /t/ or /Ø/) nearly always coincides, deviating only in classes IX and X. Consequently, although two different rules are required, most of the evaluative steps can be shared. This is illustrated in (24)–(25).

(24) `VERB:`

```
    ...
    <suffix inf> == INF_NODE1:<"<segment stem_final
     juss>"> a q
    <suffix ger> == GER_NODE1:<"<segment stem_final
     juss>"> e n ...
```

(25) `GER_NODE1:`

```
    <l> == INF_NODE1:<"<segment stem_final juss>">
    <> == GER_NODE2:<"<class stem>">.
  GER_NODE2:
    <> == INF_NODE1:<"<segment stem_final juss>">
    <strengthening> == FORM_DENTAL:<fortis>.
```

Starting at the VERB node in (24), the infinitive and gerund are directed towards different evaluative paths, but the rule sequence for the gerund (25) quickly rejoins that of the infinitive at INF_NODE1. Thus, partial implicature corresponds here to partial morphological identity.

4.4 *Varieties of lexical specification*

In the vast majority of cases (*c.* 98% of the lexicon), information about the stem-final syllable and the stem gradation pattern is enough to generate the correct distribution of suffix allomorphs. Lexical specification is still possible

33

though, via the `class suffix` feature, which can have the lexically speci-fied value `null`. However, because so few rules make reference to this feature – the present 3rd person, the infinitive and, by extension, the gerund – and because reference to the `class suffix` feature is embedded within other suffix assignment rules, the opportunities for lexical specifi-cation to have any surface manifestation are limited. In addition to these systematic exceptions, there are three lexemes that show highly aberrant suffixation patterns which the regular system of rules cannot accommodate. In these cases the forms are directly specified in the lexical entry, cir-cumventing the rule system. Thus, with respect to overt specifcation of the suffixes, there are three types of lexical entry, exemplified by the (abbrevi-ated) entries given in examples (26)–(28): no lexical specification, as in (26), suffixation based on `class suffix`, as in (27), and directly specified ab-errant suffixation, as in (28).[22]

(26) PÜGÄMÄ_IV_76:
    `< > == VERB`
    `<class stem> == weakening a`
    `<segment stem_final impers prs> == $diphthong`
    `<stress stem_final impers prs> == stressed.`

(27) SAAMA_II_78:
    `< > == VERB`
    `<segment stem_final> == $long_vowel`
    `<stress stem_final impers prs> == stressed`
    `<class suffix> == null.`

(28) TULÕMA_III_79:
    `< > == VERB`
    `<class stem> == geminating a`
    `<segment stem_final juss> == l`
    `<segment stem_final impers> == l`
    `<stress stem_final juss> == stressed`
    `<stress stem_final impers prs> == stressed`
    `<suffix prs 3> == FORM_PRS:<null>`
    `<suffix impers ptcp> == t.`

5. CONCLUSION

Viewed as an isolated system, the suffix classes of Võro verbs look as if they lack clear organizing principles. Although the actual number of allomorphs

---

[22] See Appendix C for the format of the lexical entry.

in competition with each other is very small (two or three), they combine so freely with each other that we would need to posit nearly two dozen inflection classes to account for their paradigmatic distribution – classes which would consist largely in the mere listing of forms, without implicational structure. However, as demonstrated above in Section 3, the surface suffix classes are largely determined by other components of the inflectional paradigm, so that implicational structure is found across these structures, and not between suffixes. In part allomorph choice is determined by the phonological properties of the stem, though of greater interest to morphological theory are the cases where stem alternation patterns determine or influence suffix allomorphy. Within this there is still room for lexical specification of suffix allomorphs – inflection classes in the strict sense – but its role is highly restricted. In this respect the Võro system is by no means unique; what is striking is rather the relative equilibrium of the various components, which makes it particularly hard to distinguish between, say, phonologically- and morphologically-based inflection class distinctions. For this reason, the formal model in Section 4 is an ecletic representation of inflectional rules, combining the various elements in a unified analysis, in which the varying contributions of phonological, morphological and lexical information can be tracked. As such, it is a concrete implementation of the thoughts of various researchers (see Section 1) who stress the need to treat inflectional paradigms as composed of interrelated networks.

This model also provides an answer to the question of what needs to be known, at the level of the individual lexeme, in order to produce the correct suffix allomorphs. On the assumption that stem phonology is necessarily part of the lexical entry, along with the stem alternation pattern (since it is realized through stem phonology), the answer is: not very much. At most 2% of the verbal lexicon needs to be lexically specified as having non-default suffix allomorphs. How these allomorphs are actually distributed in the paradigm is a consequence of other properties (stem phonology and stem gradation pattern), and need not be specified in the lexical entry. This is not to say that the system as a whole is simple. The network of implicational rules is dense, and the stem alternation patterns which are so crucial to determining suffix allomorphy themselves describe a system of no mean complexity.

Crucially, though, we can derive the complexity of the suffix classes from other aspects of the lexical entry. But is this necessarily the case? The alternative would be a language in which we find the same suffix system as in Võro verbs, but no other manifestations of inflection (and hence, no conditioning environments), so that the distribution of the suffixes would have to be lexically specified on a lexeme-by-lexeme basis. But to the best of my knowledge, unconditioned distributional complexity of this degree, if not impossible, seems at least to be

35

rare.[23] This could be because of inherent constraints, along the lines of the Paradigm Economy Principle or the Low Entropy Conjecture (Malouf & Ackerman 2010), which are based on the idea that inflectional systems tend to limit complexity. It could also simply be the by-product of how inflectional allomorphy comes about in the first place, in as much as there must have been a conditioning environment at some diachronic point; while such environments may be obscured over time, it would be surprising to see them effaced altogether. Any resolution to this question will depend upon further study of complex inflectional systems.

APPENDIX A

More on stem gradation

In Section 3.2 a reduced portrayal of the stem gradation patterns was given, on the argument that further details were not relevant for predicting suffix allomorphy. A fuller picture is given here, in terms of the full set of reference forms discussed in the paper, shown in Table A1; see also Iva's (2010) account in English of stem gradation. This shows that there are gradation alternations besides those discussed above; they have been omitted because they are not demonstrably predictive of suffix allomorphy. For example, the present participle always displays strong grade, and the impersonal forms always display weak grade. The only exception to this are the Q3 invariant stems such as `nälgümä 'starve'; thus the presence of a gradation alternation between these two values has as good as no predictive value. With Q1 stems, infinitive -Aq and gerund -En cooccurs with gemination/fortition of the final stem consonant, cross-cutting other gradation distinctions (compare püka-q, sipa-q, kossi-q); in the case of püka-q, gemination/fortition is applied to what is, in its paradigmatic context, already a strong stem (püga). This COULD be construed as a prediction that goes from stem to suffix, which would require adding at least one other gradation class, to account for Q1 stems that lack gemination/fortition for these values. But the reverse prediction would not involve adding additional classes, and fits in with larger generalizations. If gemination/fortition were construed as a morphophonological effect of -Aq/-En suffixation, this would cover all instances, and parallels the effects these suffixes appear to induce with diphthong-final stems, e.g. `vao-ma 'sink' ∼ vaio-q.

---

[23] In some Chinantecan languages, the tonal paradigms marking subject person-number and TAM yield as many as 100 classes, all derived through various combinations of a small set (about a dozen) of tone distinctions in a twelve-cell paradigm (see e.g. Finkel & Stump 2009, Palancar forthcoming). These classes are not readily derivable from any other aspect of the paradigm, so that these are among the most extreme examples of distributional complexity I am aware of. But tone is always realized as part of a syllable that naturally has its own phonological properties. It could be that this dependence plays a role in facilitating these patterns.

37

	Weakening		Strengthening	Geminating		Invariant	
	type a 'shear'	type b 'be born'	'revile'	type a 'tire'	type b 'scurry'	Q1 stem 'grow'	Q3 stem 'starve'
PRS 3SG	**pügä**	sünnü-s	**`sõima-s**	väsü-s	**sipa**	kosi	`nälgü-s
PST I	`pöi	**sündü**	**`sõim-ssi**	**`vässü**	siba-si	kosõ	`nälgü
INF	**pükä-q**	**sündü-däq**	sõima-daq	**vässü-q**	**sipa-q**	**kossi-q**	`nälgü-däq
GER	**pükä-n**	**sündü-den**	sõima-tõn	**vässü-n**	**sipa-n**	**kossi-n**	`nälgü-den
JUSS	**pügä-guq**	**`sündü-guq**	**`sõima-kuq**	väsü-guq	siba-guq	kosi-guq	`nälgü-guq
IMPERS PRS 3SG	pöet-as	sünnü-täs	sõima-tas	väsü-täs	siba-tas	kosi-tas	`nälgü-täs
IMPERS PTCP PST SG	`pöe-t	sünnü-t	sõima-t	väsü-t	siba-t	kosi-t	`nälgü-t
PTCP PRS	**pükä-v**	**`sündü-v**	**`sõima-v**	**vässü-v**	**sipa-v**	**kossi-v**	`nälgü-v

Table A1
Stem gradation patterns, strong grade stems shown in boldface; gradation alternations discussed in this appendix are shaded.

Long vowel-final stems display a unique pattern of stem alternation (see Table A2), though it is not clear that there is any compelling reason to treat these as part of the gradation system. First, there is an alternation between *VV* and *Vj* or *Vvv* (where *V* is a back vowel). The *Vj/Vvv* alternants are correlated with vowel-initial suffixes (on the assumption that the past 1st person suffix is, at some level, really *V*), so this may best be seen as a concommittant of suffixation (see Section 3.1). With the *VV* forms, the alternation between strong and weak grade forms can, at least in most cases, be derived from non-lexical properties. First, phonology dictates that monosyllabic forms are necessarily Q3. Second, as noted above, the present participle has strong grade across the whole lexicon. However, even if we choose to treat this as a type of gradation alternation, it would not afford any predictions beyond those already deduceable from stem phonology.

	'stay'
PRS 3SG	ˋjää-s
PST I	ˋjäi
INF	jäi-äq
GER	jäi-en
JUSS	jää-guq
IMPERS PRS 3SG	jäi-äs
IMPERS PTCP PST SG	ˋjää-d
PTCP PRS	ˋjää-vä

Table A2
Stem alternations of long vowel-final stems.

Finally, there is an additional gradation alternation found with stems that end in a fortis dental stop or affricate (*t*, *ts*), which regularly display lenition in the 1st person present and past, e.g. 1st person *kergüdä* 'raise˙' vs. second person *kergütä*, 3SG *kergütä-s*. Because this alternation applies across the whole lexicon, it does not have any predictive value beyond what is already provided stem phonology.

APPENDIX B

Rules for Võro verb suffixation

The rules below generate the stem-final and suffixes for those parts of the paradigm that display suffix allomorphy (see Table 6 in the main text); only a fragment of the extensive mutually implicative impersonal paradigm is shown here, namely the present 3rd person singular and plural.

```
VERB:
   <> ==
   <end> == "<segment stem_final>" "<suffix>"
   <end impers prs 3> == "<segment stem_final juss>" <suffix
    impers>

FORM_PRS:<>
   <segment stem_final> == $vowel
   <stress stem_final> == unstressed
   <suffix prs 3> == FORM_PRS:<"index">
   <index> == PRS_NODE1:<"<segment stem_final prs>">
   <suffix pst> == PST_NODE1:<"<segment stem_final juss>">
    <termination pst>
   <suffix inf> == INF_NODE1:<"<segment stem_final juss>"> a q
   <suffix ger> == GER_NODE1:<"<segment stem_final juss>"> e n
   <suffix juss> == JUSS_NODE1:<"<segment stem_final juss>"> u q
   <suffix imp pl> == JUSS_NODE1:<"<segment stem_final
    juss>">e q
   <suffix impers> == IMPERS_NODE1:<"<segment stem_final
    impers prs>"> a
   <suffix impers ptcp pst sg> == IMPERS_PTCP:<"<segment
    stem_final impers ptcp>">
   <suffix ptcp prs> == v FORM_A:<"<segment stem_final ptcp
    prs>">
   <termination pst 2> == t
   <termination pst 3 pl> == q.
```

Evaluative nodes

```
PRS_NODE1:
   <$consonant> == null
   <$diphthong> ==
   <> == PRS_NODE2:<"<class stem>">.

PRS_NODE2:
   <weakening a> == null
   <geminating b> == null
   <> == "<class suffix>".

PST_NODE1:
   <l> == FORM_PST
   <$diphthong> == FORM_PST
   <$long_vowel> == FORM_PST
   <a> == PST_NODE2:<"<segment stem_penult ptcp prs>">
   <e> == PST_NODE3:<"<segment stem_penult ptcp prs>">
   <> == FORM_PST:<"<class stem>">.

PST_NODE2:
   <h> == FORM_PST:<strengthening>
   <> == FORM_PST:<"<class stem>">.
```

39

```
PST_NODE3:
   <n> == FORM_PST:<strengthening>
   <> == FORM_PST:<"<class stem>">.

INF_NODE1:
   <$long_vowel> == FORM_DENTAL:<null>
   <$consonant> == FORM_DENTAL:<null>
   <$diphthong> == FORM_DENTAL:<"<class suffix>">
   <l> == INF_NODE2:<"<stress stem_final juss>">
   <s> == FORM_DENTAL:<fortis>
   <a> == INF_NODE3:<"<segment stem_penult ptcp prs>">
   <> == INF_NODE4:<"<class stem>">.

INF_NODE2:
   <stressed> == FORM_DENTAL:<null>
   <> == FORM_DENTAL.

INF_NODE3:
   <t> == FORM_DENTAL:<null>
   <> == INF_NODE4:<"<class stem>">.

INF_NODE4:
   <weakening a> == FORM_DENTAL:<null>
   <weakening b> == FORM_DENTAL
   <strengthening> == FORM_DENTAL
   <geminating> == FORM_DENTAL:<null>
   <> == FORM_DENTAL:<"<class suffix>">.

GER_NODE1:
   <l> == INF_NODE1:<"<segment stem_final juss>">
   <> == GER_NODE2:<"<class stem>">.

GER_NODE2:
   <> == INF_NODE1:<"<segment stem_final juss>">
   <strengthening> == FORM_DENTAL:<fortis>.

JUSS_NODE1:
   <> == FORM_VELAR
   <$consonant> == FORM_VELAR:<strengthening>
   <$vowel> == FORM_VELAR:<"<class stem>">
   <e> == JUSS_NODE2:<"<segment stem_penult ptcp prs>">.

JUSS_NODE2:
   <n> == FORM_VELAR
   <> == FORM_VELAR:<"<class stem>">.

IMPERS_NODE1:
   <l> == IMPERS_NODE2:<stressed>
   <> == IMPERS_NODE2
   <$long_vowel> == FORM_DENTAL:<null>
   <$vowel> == IMPERS_NODE2:<"<stress stem_final impers
    prs>">.
```

40

```
IMPERS_NODE2:
  <stressed> == FORM_DENTAL
  <> == FORM_DENTAL:<fortis>.

IMPERS_PTCP:
  <l> == FORM_DENTAL
  <$long_vowel> == FORM_DENTAL
  <> == FORM_DENTAL:<fortis>.
```

Form nodes

```
FORM_PRS:
  <> == CLASS_DEFAULT:<>
  <null> == CLASS_NULL:<>.

CLASS_DEFAULT:
  <sg> == s
  <pl> == s e d.

CLASS_NULL:
  <sg> ==
  <pl> == v a d.

FORM_PST:
  <strengthening> == s i
  <geminating b> == <strengthening>
  <> == vowel.

FORM_DENTAL:
  <null> ==
  <fortis> == t
  <> == d.

FORM_VELAR:
  <strengthening> == k
  <> == g.

FORM_A:
  <$vowel> ==
  <> == a.
```

APPENDIX C

Exemplary lexicon of Võro verbs

This gives Iva's (2007) exemplary verb types, with the necessary minimum of information about stem alternations and gradation to predict suffixation according to the rules in Appendix B. The names of each entry follow the format [citation form of exemplary verb]+[Roman numeral classification of suffix pattern according to scheme in figure (3) of the paper]+[Iva's (2007) classificatory number]. In the the lexicon, instances of lexical specification, either by the class feature (indicating the exceptional inflection class) or by direct lexical specification of the form, are highlighted through shading.

41

```
SIBAMA_I_88:
   <> = = VERB
   <class stem> = = geminating b.

SÜÜMÄ_II_77:
   <> = = VERB
   <segment stem_final> = = $long_vowel
   <segment stem_final imperf> = = $vowel
   <stress stem_final impers prs> = = stressed
   <class suffix> = = null.

SAAMA_II_78:
   <> = = VERB
   <segment stem_final> = = $long_vowel
   <stress stem_final impers prs> = = stressed
   <class suffix> = = null.

TULÕMA_III_79:
   <> = = VERB
   <class stem> = = geminating a
   <segment stem_final juss> = = l
   <segment stem_final impers> = = l
   <stress stem_final juss> = = stressed
   <stress stem_final impers prs> = = stressed
   <suffix prs 3> = = FORM_PRS:<null>
   <suffix impers ptcp> = = t.

PÜGÄMÄ_IV_76:
   <> = = VERB
   <class stem> = = weakening a
   <segment stem_final impers prs> = = $diphthong
   <stress stem_final impers prs> = = stressed.

VALAMA_IV_75:
   <> = = VERB
   <class suffix> = = null.

KOSIMA_IV_87:
   <> = = VERB
   <class suffix> = = null.

RAPAHUTMA_V_80:
   <> = = VERB
   <segment stem_final> = = t
   <segment stem_final impers> = = $vowel.

LASKMA_V_84:
   <> = = VERB
   <class stem> = = weakening a
   <segment stem_final> = = k
   <segment stem_final impers> = = s
   <stress stem_final impers prs> = = stressed.
```

42

```
PÜÜDMÄ_V_82:
   <> == VERB
   <class stem> == weakening a
   <segment stem_final> == $consonant.

NÕSTMA_V_83:
   <> == VERB
   <class stem> == weakening a
   <segment stem_final> == $consonant
   <segment stem_final impers> == $vowel.

HIRNMA_V_85:
   <> == VERB
   <class stem> == weakening a
   <segment stem_final> == $consonant
   <segment stem_final impers> == $vowel.

RISTMÄ_V_86:
   <> == VERB
   <class stem> == weakening a
   <segment stem_final> == $consonant
   <segment stem_final impers> == $vowel.

ISTMA_2_V_86:
   <> == VERB
   <class stem> == weakening a
   <segment stem_final> == $consonant
   <segment stem_final impers> == $vowel.

TEGEMÄ_VI_81:*
   <> == VERB
   <class stem> == weakening a
   <stress stem_final impers prs> == stressed
   <segment stem_final prs> == $vowel
   <segment stem_final pst> == $vowel
   <segment stem_final> == s
   <end> == $vowel <suffix>
   <suffix impers ptcp> == tt.

HARINÕMA_2_VII_67:
   <> == VERB
   <segment stem_final> == e
   <segment stem_penult ptcp prs> == n
   <stress stem_final impers prs> == stressed.
```

---

[*] The lexical entry reflects the fact that the suffixes are those that would be expected in case of a stem-final /s/, but the actual stem-final segment is a vowel.

43

```
HARINÕMA_1_VIII_66:
   <> == VERB
   <segment stem_final> == e
   <segment stem_penult ptcp prs> == n.

SIBAHAMA_VIII_65:
   <> == VERB
   <segment stem_final> == a
   <segment stem_penult ptcp prs> == h.

ARGNÕMA_IX_62:
   <> == VERB
   <class stem> == strengthening
   <segment stem_final> == e
   <segment stem_penult ptcp prs> == n.

SÕIMAMA_X_63:
   <> == VERB
   <class stem> == strengthening.

PUHKSAMA_XI_64:
   <> == VERB
   <class stem> == strengthening
   <segment stem_final juss> == s
   <segment stem_final impers> == s.

JÄÄMÄ_XII_53:
   <> == VERB
   <segment stem_final> == $long_vowel
   <stress stem_final impers prs> == stressed.

EHITE(L)LEMÄ_XIII_52:
   <> == VERB
   <segment stem_final juss> == l
   <segment stem_final impers> == l
   <stress stem_final impers prs> == stressed
   <stress stem_final inf> == stressed
   <stress stem_final juss> == stressed.

HIRNAHTAMA_XIV_58:
   <> == VERB
   <segment stem_final> == a
   <segment stem_penult> == t
   <stress stem_final impers prs> == stressed.

KERGÜTÄMÄ_XIV_59:
   <> == VERB
   <segment stem_final> == a
   <segment stem_penult> == t
   <stress stem_final impers prs> == stressed.
```

44

```
KAOTAMA_XIV_59:
  <> == VERB
  <segment stem_final> == a
  <segment stem_penult> == t
  <stress stem_final impers prs> == stressed.

ELÄMÄ_XV_48:
  <> == VERB
  <class stem> == geminating a.

KASUMA_XV_48:
  <> == VERB
  <class stem> == geminating a.

VASÜMA_XV_74:
  <> == VERB
  <class stem> == geminating a.

VAOMA_2_XVI_73:
  <> == VERB
  <segment stem_final> == $diphthong
  <stress stem_final impers prs> == stressed
  <class suffix> == null.

MANITSÕMA_1_XVII_60:
  <> == VERB
  <segment stem_final juss> == $consonant
  <stress stem_final impers prs> == stressed.

ISTMA_1_XVIII_54:
  <> == VERB
  <class stem> == weakening b
  <segment stem_final juss> == $consonant
  <segment stem_final ptcp prs> == $consonant.

LÕPMA_XVIII_49:
  <> == VERB
  <class stem> == weakening b
  <segment stem_final juss> == $consonant
  <segment stem_final ptcp prs> == $consonant.

KÕNÕLÕMA_XIX_55:
  <> == VERB
  <segment stem_final juss> == l
  <segment stem_final impers ptcp> == l.

HIIBÕLÕMA_XIX_56:
  <> == VERB
  <class stem> == strengthening
  <segment stem_final juss> == l
  <segment stem_final impers> == l.
```

45

```
TAPLÕMA_XIX_57:
  <> == VERB
  <class stem> == strengthening
  <segment stem_final juss> == l
  <segment stem_final impers> == l.

ERÄTÜMÄ_XX_71:
  <> == VERB
  <stress stem_final impers prs> == stressed.

UNÕH(T)UMA_XX_70:
  <> == VERB
  <stress stem_final impers prs> == stressed.

SÜNDÜMÄ_XXI_69:
  <> == VERB
  <class stem> == weakening b.

MANITSÕMA_2_XXI_61:
  <> == VERB.

NÄLGÜMÄ_XXI_68:
  <> == VERB.

VAOMA_1_XXII_72:
  <> == VERB
  <segment stem_final> == $diphthong
  <stress stem_final impers prs> == stressed.

HELISEMÄ_XXIII_50:
  <> == VERB
  <segment stem_final> == s
  <segment stem_final prs> == e.

NÕSÕMA_XXIII_51:
  <> == VERB
  <class stem> == geminating a
  <segment stem_final juss> == s
  <segment stem_final impers> == s
  <stress stem_final impers prs> == stressed.
```

REFERENCES

Anderson, Stephen R. 2008. Phonologically conditioned allomorphy in the morphology of Surmiran (Rumantsch). *Word Structure* 1, 109–34.

Baayen, R. Harald, James McQueen, Ton Dijkstra & Rob Schreuder. 2003. Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R. Harald Baayen & Rob Schreuder (eds.), *Morphological structure in language processing*, 355–390. Berlin: Mouton.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42, 531–573.

Blevins, James P. 2007. Conjugation classes in Estonian. *Linguistica Uralica* XLIII.4, 250–267.

Blevins, James P., Farrell Ackerman & Rob Malouf. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In Blevins & Blevins (eds.), 54–82.

Blevins, James P. & Juliette Blevins (eds.). 2009. *Analogy in grammar: Form and acquisition*. Oxford: Oxford University Press.

Brown, Dunstan, Greville G. Corbett, Norman Fraser, Andrew Hippisley & Alan Timberlake. 1996. Russian noun stress and Network Morphology. *Linguistics* 34, 53–107.

Brown, Dunstan & Andrew Hippisley. 1994. Conflict in Russian genitive plural assignment: A solution represented in DATR. *Journal of Slavic Linguistics* 2.1, 48–76.

Brown, Dunstan & Andrew Hippisley. 2012. *Network Morphology: A defaults-based theory of word structure* (Cambridge Studies in Linguistics 133). Cambridge: Cambridge University Press.

Cameron-Faulkner, Thea & Andrew Carstairs-McCarthy. 2000. Stem alternants as morphological signata: Evidence from blur avoidance in Polish nouns. *Natural Language & Linguistic Theory* 18, 813–835.

Carstairs[-McCarthy], Andrew. 1983. Paradigm economy. *Journal of Linguistics* 19, 115–28.

Carstairs[-McCarthy], Andrew. 1988. Some implications of phonologically conditioned suppletion. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 1988*, 68–94.

Corbett, Greville G. & Norman Fraser. 1993. Network Morphology: A DATR account of Russian nominal inflection. *Journal of Linguistics* 29, 113–142.

Evans, Roger & Gerald Gazdar. 1996. DATR: A language for lexical knowledge representation. *Computational Linguistics* 22.2, 167–216.

Finkel, Raphael & Gregory T. Stump. 2007. Principal parts and morphological typology. *Morphology* 17.1, 39–75.

Finkel, Raphael & Gregory T. Stump. 2009. Principal parts and degrees of paradigmatic transparency. In Blevins & Blevins (eds.), 13–53.

Iva, Sulev. 2007. *Võru kirjakeele sõnamuutmissüsteem*. Ph.D. thesis, University of Tartu. [dspace.utlib.ee/dspace/bitstream/10062/4860/1/iva_sulev.pdf]

Iva, Sulev. 2010. Grade alternation in Võro South Estonian. *Linguistica Uralica* XLVI 2010.3, 161–174.

Kaalep, Heiki-Jaan & Kadri Muischnek. 2002. *Eesti kirjakeele sagedussõnastik*. Tartu: TÜ Kirjastus.

Keem, Hella. 1997. *Võru keel*. Tallinn: Eesti Teaduste Akadeemia.

Larsen, Robert E. 1977. Orokaiva grammar. Ms., Summer Institute of Linguistics. http://www.sil.org/pacific/png/pubs/928474531028/Orokaiva_Grammar.pdf (accessed 30 March 2010).

Malouf, Rob & Farrell Ackerman. 2010. Paradigms: The low entropy conjecture. Presented at Workshop on Morphology and Formal Grammar, Université Paris Diderot, 8 July.

Müller, Gereon. 2007. Notes on paradigm economy. *Morphology* 17.1, 1–38.

Palancar, Enrique L. Forthcoming. Revisiting the complexity of the Chinantecan verb conjugation classes. In Jean-Léo Léonard & Alain Kihm (eds.), *Issues in Meso-American morphology*. Paris: Michel Houdiard. [To appear in 2013]

Stump, Gregory T. & Raphael Finkel. 2010. Principal parts and morphological analysis. Presented at Workshop 'Morphological Complexity: Implications for the Theory of Language', Harvard University, 22 January.

Viitso, Tiit-Rein. 2003. Rise and development of the Estonian language. In Mati Erelt (ed.), *Estonian language* (Linguistica Uralica supplementary series 1), 120–230.

Viks, Ülle. 1992. *Väike võrmisõnastik*, vol. 1. Tallinn: Eesti Teaduste Akadeemia, Keele ja Kirjanduse Instituut.

Wurzel, Wolfgang U. 1984. *Flexionsmorphologie und Natürlichkeit*. Berlin: Akademie Verlag.

Author's address: Surrey Morphology Group [I1], University of Surrey,
Guildford, Surrey, GU2 7XH, UK
m.baerman@surrey.ac.uk