Ossetic verb morphology in LrFG

Oleg Belyaev

Lomonosov Moscow State University
Institute of Linguistics RAS

23 July 2023

LFG Conference
Rochester University
The descriptive aim of this talk is to formalize Ossetic verb morphology in LFG, especially focusing on the following issues:

- present vs. past stem alternations;
- transitivity marking in past stem forms;
- “three-stem” intransitive verbs.

In a wider perspective, I want to discuss the following topics:

- compatibility of LrFG with “classic” LFG assumptions;
- the handling of morphomic stems in LrFG, and the existence of “autonomous” morphology in general.
Like in other modern Iranian languages (see Dashti & Asudeh, 23.07.23), Ossetic verbs use two stems, traditionally called “present” and “past”:

- $kən$- / $ko(n)d$- ‘do’
- $səw$- / $səd$- ‘go’
- $nəmaj$- / $nəmad$- ‘count’
- $šuz$- / $səʁd$- ‘burn’
- $əft$- / $əftəd$- ‘increase’
- ...

The model of stem derivation is, in general, lexical information (neither stem is predictable from the other).

The only pattern that never involves any vocalic or consonantal alternations is the $-əd$ pattern (cf. Persian $-id$).
In Modern Persian, the present and past endings are largely identical (except for 3sg), and the stem can, in most contexts, be associated with the corresponding TAM feature (nonpast / past):

- **present stem:**
  - present *mi-kon-am* ‘I am doing’,
  - subjunctive *be-kon-am* ‘I would do’;
- **past stem:**
  - imperfect *mi-kard-am* ‘I was doing’,
  - aorist *kard-am* ‘I did’.

But Modern Persian is somewhat of an exception: the distribution of stems often has a less straightforward motivation.

Cf. Derbend Muslim Tat (Jalqan):

- **“past” stem:**
  - present *mi-sæxt-am* ‘I am doing’,
  - future *sæxt-eni-m* ‘I will do’,
  - aorist *sæxt-üm* ‘I did’,
  - “past eventual” *mi-sæxt-üm* ‘I would have done’;
- **“present” stem:**
  - imperative *sæs* ‘do!’,
  - “eventual” *mi-sæs-üm* ‘I would do’,
  - subjunctive *sæs-üm* ‘that I do’.
Functions of Ossetic stems

- Ossetic is less radical, but stems cannot be assigned any temporal value.
- Present stem:
  - temporal forms: present səw-ən ‘I go’, future səw-zən-ən ‘I will go’;
  - modal forms: subjunctive səw-on ‘I would go’, optative səw-in ‘I would like to go’, imperative su ‘go!’;
  - present participles: səw-əg, səw-ag, səw-gə;
  - future participle səw-inəg;
  - infinitive səw-ən;
  - destinative participle səw-ən ‘for walking’.
- Past stem:
  - past səd-tən ‘I went’;
  - counterfactual səd-aın ‘I would have gone’;
  - “past” participle (+ nominalization) səd.
- Each tense-mood paradigm has its own set of endings
  - ⇒ stems are redundant,
  - they are morphomic (aronoff1976); the stem is not an f-structure feature.
Aspect

- Stems never display aspect.
- This is the function of preverbs (Slavic-style system, Bybee and Dahl 1989):
  - *kod-t-on* ‘I did / was doing’ (ipfv.) → *ba-kod-ton* ‘I did’ (pfv.)
  - *fəšt-on* ‘I wrote / was writing’ (ipfv.) → *nə-ffəšt-on* ‘I wrote’ (pfv.)
- There is usually a “default” preverb lexically associated with a verb stem, but other preverbs may convey additional aspectual meanings (Tatevosov 2019)
- Preverbs clearly attach above the level of V, cf. complex verbs:
  - *ləg kɜn-* ‘cut’ (ipfv.) → *a-ləg kɜn-* ‘cut’ (pfv.)
- Therefore, their behaviour is not directly relevant and will not be treated here
- I ( provisionally ) assume that aspect is expressed as the f-structure feature ASPECT with features PFV / IPFV
- Since each tense-mood category is identified by its unique set of endings, I analyze it as a holistic f-structure feature VFORM.
Ossetic verb morphology

Two stems

\[ \text{ba-kod-t-ain} \]
\[ \text{PV-do-tr-cntrf.1sg} \]

\[ \text{kən-а} \]
\[ \text{do-sbjv.3sg} \]

\[ \text{s3w-ə} \]
\[ \text{go-prs.3sg} \]
Transitivity

- Ossetic also has morphological transitivity marking in finite forms derived from the past stem.
- An extra -t- appears between the stem and the ending:
  - it may surface as gemination (after vowels and sonorants): nad-t-ain → nat:ain ‘I would have beaten’,
  - as devoicing: žayd-t-on → žaxton ‘I said’ (cf. ptcp. žayd),
  - or as nothing: fešt-t-on → fešton ‘I wrote’.
- Labile “minimal pair”: counterfactual intr. šayd-aid ‘it would have burnt’, šeχt-aid (*šayd-t-aid) ‘s/he would have burnt smth.’
- Past tense forms use a separate set of endings: šayd-i ‘it burnt’, šeχt-a ‘s/he burned’
  - note, however, that the transitive past endings are all but identical to the subjunctive endings (except 1PL),
  - synchronically, I treat them as two distinct set because the meanings are too distant (and based on different stems).
Overall, this tends to correlate with syntactic transitivity, but not exactly (Vydrin 2022):

- $k\text{ɜw-} / k^w\text{əd-} ‘cry’: tr. $k^w\text{ət}:a ‘cried’;
- $\text{wom-} / \text{womd-} ‘vomit’: tr. $\text{womt}:a ‘vomited’;
- $\text{žɣor-} / \text{žɣord-} ‘run’: tr. $\text{žɣort}:a ‘ran’, etc.

Around 30 intransitive verbs with transitive endings

- sometimes with close semantic counterparts: $\text{wajən} ‘run’ (intr.)

The converse also happens, although less frequently: $\text{məʃ-} / \text{məʃəd-} ‘remember’ (intr.) takes a DO.

List of “exceptions” similar to split S marking in other Iranian languages (Chistyakova 2023).

Morphological transitivity has no known syntactic effects

- an inflectional class of verbs.
Some verbs have three stems rather than two (Abaev 1959, 55–56):

- prs. *waš-* / pst. *waš-əd* / ptcp. *wašt* ‘cry’ (of animal);

All these verbs have two properties in common:

- they are (morphologically) intransitive;
- the past stem uses the “regular” suffix -əd while the participle uses the generally irregular -d/-t.

This pattern requires an explanation: why do transitive verbs never have a distinct participial stem?
Illustrations

- inflection of the intransitive verb $xʷəšš- / xʷəšš-əd- / xʷəšt$ ‘sleep’

<table>
<thead>
<tr>
<th></th>
<th>present</th>
<th>preterite</th>
<th>counterfactual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sg.</td>
<td>pl.</td>
<td>sg.</td>
</tr>
<tr>
<td>1</td>
<td>$xʷəšš-ən$</td>
<td>$xʷəšš-əm$</td>
<td>$xʷəšš-əd-tən$</td>
</tr>
<tr>
<td>2</td>
<td>$xʷəšš-əš$</td>
<td>$xʷəšš-ut$</td>
<td>$xʷəšš-əd-tə$</td>
</tr>
<tr>
<td>3</td>
<td>$xʷəšš-ə$</td>
<td>$xʷəšš-əns$</td>
<td>$xʷəšš-əd(-i(š))$</td>
</tr>
</tbody>
</table>

participle: $xʷəšt$

- inflection of the transitive verb $šur- / šərd$ ‘chase’

<table>
<thead>
<tr>
<th></th>
<th>present</th>
<th>preterite</th>
<th>counterfactual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sg.</td>
<td>pl.</td>
<td>sg.</td>
</tr>
<tr>
<td>1</td>
<td>$šur-ən$</td>
<td>$šur-əm$</td>
<td>$šərd-t-on$</td>
</tr>
<tr>
<td>2</td>
<td>$šur-əš$</td>
<td>$šur-ut$</td>
<td>$šərd-t-aj$</td>
</tr>
<tr>
<td>3</td>
<td>$šur-ə$</td>
<td>$šur-əns$</td>
<td>$šərd-t-a$</td>
</tr>
</tbody>
</table>

participle: $šərd$
The role of the transitive marker -t-:
- why does it appear with otherwise identical endings (Counterfactual) depending on transitivity?

The status of three-stem verbs.
General idea

- Recall that 3-stem verbs have a limited distribution:
  - intransitive
  - past stem in -əd

- The status of the “past stem” is different for transitive and intransitive verbs
  - **intransitive** all verbs have a past stem, which also acts as a participle; some verbs also have a dedicated, irregular participle form;
  - **transitive** there is no past stem, only the participle; -t- acts to verbalize the participle in order to allow it to take tense-mood endings.
The order of stem formation is effectively inverse depending on inflection class:

- **transitive**: the lexicon stores the present stem and the participle; to form finite forms, a new stem must be produced from the participle using the suffix -t;

- **intransitive**: the lexicon stores the present and past stems; the participle has no separate exponent and is identical to the past stem by default (Pāṇini’s Principle, Anderson 1969), but some verbs also store a separate participial stem alongside a regular past stem.
L_{RG} is a radical departure from LFG’s understanding of Lexical Integrity: a kind of hybrid between DM and LFG, or a non-transformational (representational) variant of DM.

Core tenets:

- the leaves of the c-structure tree are \textit{f-descriptions} (+ meaning constructors) rather than words;
- these f-descriptions are realized by Vocabulary Items (VI);
- the way that VIs realize syntactic information is determined by the function $\nu$ that maps c-structure to v(ocabulary)-structure;
- among several VIs that can realize an f-description, the most informative (specific) one is selected.
An example of a VI: English verb *see*

\[
< [\sqrt], \Phi\{(\uparrow \text{PRED}) = \text{‘see’}\}, \{\text{see} : (\uparrow \text{OBJ})_\sigma \rightarrow (\uparrow \text{SUBJ})_\sigma \rightarrow \uparrow_\sigma\} > \nu \rightarrow \text{see}
\]

VIs can span several c-structure nodes. For example, assuming that the stem *znaj*- of the Russian verb ‘know’ is not acategorial, we can define it as a span of the root and “little ν”:

\[
< [\sqrt, \nu], \Phi\{(\uparrow \text{PRED}) = \text{‘know’}\}, \{\text{know} : (\uparrow \text{OBJ})_\sigma \rightarrow (\uparrow \text{SUBJ})_\sigma \rightarrow \uparrow_\sigma\} > \nu \rightarrow \text{znaj-}
\]

Another type of spanning is called **Pac-Man spanning**: a head may be realized by the same VI that realizes an adjacent head if no suitable dedicated candidate is available.

Example (Asudeh, Bögel, and Siddiqi 2022): English *-en* is only available after obstruents, hence *to reddens*, *to red*. When it is not available, the verbal stem undergoes Pac-Man spanning: *to orangel*.
More detailed v-structure

\[ \langle \ldots, \ldots, \ldots \rangle \xrightarrow{\nu} \]

- PHON(OGICAL)
- REP(RESENTATION)
- P(ROSODIC)FRAME
- P(ROSODIC)LEVEL \(1 \mid 2\)
- DEP(ENDENCE) \(\{LT, RT\}\)
- CLASS \(\{\text{inflectional classes}\}\)
- TYPE \{VERBAL | NOMINAL | ADJECTIVAL\}
- IDENT(ITY) AUNT | NIECE
- HOST

- \(\nu\)-S
  - PHON.REP \(\ldots\)
  - PFRAME \(\ldots\)
  - PLEVEL \(\ldots\)
  - DEP \(\ldots\)
  - CLASS \(\ldots\)
  - TYPE \(\ldots\)
LrFG and lexical integrity

- As observed in Melchin, Asudeh, and Siddiqi (2020), LrFG contradicts the definition of LI in Bresnan et al. (2016):
  
  *Morphologically complete words are leaves of the constituent-structure tree and each leaf corresponds to one and only one constituent-structure node. (Bresnan et al. 2016, 92)*

- However, it does not necessarily contradict the more lax definition in Bresnan and Mchombo (1995):
  
  *Words are built out of different structural elements and by different principles of composition than syntactic phrases. (Bresnan and Mchombo 1995, 181)*

- Syntactic phrases in LrFG are built out of smaller constituents, while words are constructed realizationally, based on vocabulary items. Furthermore, the degree of recursion under $X^0$ can be stipulated to be lower than above $X^0$ (as with $X^0$ adjunction in vanilla LFG, see Toivonen 2003).
In fact, since, unlike DM, LrFG does not assume a universal hierarchy of projections, there is considerable flexibility in the degree to which morphology is “syntactic”.

If affixes are kept below $X^0$ level, LrFG analyses are analogous to “sublexical rules” widely utilized in e.g. Bresnan et al. (2016), but considerably more powerful with respect to various morphological phenomena:

\[ N \rightarrow N_{\text{stem}} N_{\text{num}} N_{\text{case}} \]

When affixes appear above $X^0$, the analyses are similar to lexical sharing (Wescoat 2002; Lowe 2016; Belyaev 2021), but with the advantage that the contribution of each exponent is explicitly specified (rather than delegated to the morphological module).
Dealing with stems

- Stems could be treated as separate vocabulary items, defined as $\sqrt{\ + \ v}$ spans.
- But what about -t- in the transitive forms?
  - the same Counterfactual endings attach to both transitive and intransitive verbs,
  - thus, effectively, this is a secondary “stem” for such endings;
  - but storing the “extended” transitive stems in the lexicon is not an adequate solution.
- The pattern of 3-stem verbs also requires an explanation.
Preliminary remarks

Everything described here happens under V, outside of recursive syntax, cf. the analysis of O’dam in Everdell et al. (2021).

I use the following PS rules (all ↑ = ↓):

\[
\begin{align*}
\text{TenseP} & \rightarrow \text{vP Tense (TAM endings)} \\
\text{vP} & \rightarrow \overline{\text{v}} \text{ Pst (past stem suffix)} \\
\overline{\text{v}} & \rightarrow \sqrt{\text{v}} \text{ (present stem, always a span)} \\
\text{nP} & \rightarrow \text{vP n (participles)} \\
\text{vP} & \rightarrow \text{nP v (from participle to verb stem)}
\end{align*}
\]

I assume the model of Lovestrand and Lowe (2017), i.e. intermediate bar levels do not appear if nothing attaches on them.

I propose treating all past stems in -t/d as spans stored in the lexicon, because the alternations involved are, in the general case, unpredictable. The intransitive stems in -əd, in contrast, should be constructed in the morphology, because they form a very regular pattern.
Structure of present-stem forms

- **TenseP**
  - $\mathbf{\uparrow \downarrow}_\mathbf{V}$
  - $\mathbf{\uparrow \downarrow}_{\mathbf{T}}$
  - $\mathbf{\uparrow \downarrow}_{\mathbf{\Phi}}$
  - $\mathbf{\uparrow \downarrow}_\mathbf{\Theta}$

- **Tense**
  - $\mathbf{\uparrow \downarrow}_{\mathbf{\Phi}}$
  - $\mathbf{\uparrow \downarrow}_{\mathbf{\Theta}}$
  - $\mathbf{\uparrow \downarrow}_{\mathbf{\Theta}}$

- **PHON.REP** /šur/
- **CLASS** TR
- **STEM** PRS
- **TYPE** VERBAL

- New feature: STEM (similar to CLASS)

- $\mathbf{\Theta} \equiv \mathbf{\uparrow}_\sigma_{\mathbf{ARG1}} \mathbf{\uparrow}_\sigma_{\mathbf{ARG2}}$ (syntactic / semantic transitivity!)}
Structure of transitive past-stem forms

\[
< [\sqrt{\text{v}}, \text{n}], \Phi \left\{ (\uparrow \text{PRED}) = \text{‘chase’} \right\}, _\rightarrow ^\nu
\]

\[
\text{PHON.REP} /\text{šərd}/
\]
\[
\text{CLASS} \quad \text{TR}
\]
\[
\text{STEM} \quad \text{PST}
\]
\[
\text{TYPE} \quad \text{NOMINAL}
\]

\[
< [\text{v}], \Phi\{\}, _\rightarrow ^\nu
\]

\[
\text{PHON.REP} /\text{t}/
\]
\[
\text{DEP} \quad \text{LT}
\]
\[
\text{CLASS} \quad \text{TR}
\]
\[
\text{STEM} \quad \text{PST}
\]
\[
\text{TYPE} \quad \text{VERBAL}
\]

\[
\text{IDENT} \quad \text{NIECE}
\]
\[
\text{TYPE} \quad \text{NOMINAL}
\]
\[
\text{CLASS} \quad \text{TR}
\]
\[
\text{STEM} \quad \text{PST}
\]
Counterfactual endings are agnostic with respect to transitivity:

\[
< \text{Tense}, \Phi \left\{ \begin{array}{l}
(\uparrow \text{VFORM}) = \text{CNTRF} \\
(\uparrow \text{PERS}) = 3 \\
(\uparrow \text{NUM}) = \text{SG}
\end{array} \right\}, \_ \xrightarrow{\nu} \begin{bmatrix}
\text{PHON.REP} & /aid/ \\
\text{DEP} & \text{LT} \\
\text{STEM} & \text{PST} \\
\text{TYPE} & \text{VERBAL} \\
\text{HOST} & \begin{bmatrix}
\text{IDENT} & \text{NIECE} \\
\text{TYPE} & \text{VERBAL} \\
\text{STEM} & \text{PST}
\end{bmatrix}
\end{bmatrix}
\]
“Functionally empty” heads

- The “nominalizer” \( n \) and the “verbalizer” \( v \) in finite transitive verbs are associated with no \( f \)-descriptions.
- This is not the case for \( n \) in general:
  - \( (\uparrow \text{VFORM}) = \text{PTCP.PST} \)
  - \( (\uparrow \text{VFORM}) = \text{PTCP.PRS} \)
  - \( (\uparrow \text{VFORM}) = \text{PTCP.FUT} \)
  - ... 

- The past participle has no independent exponent, and in the case of transitive verbs is realized by the VI above (there is no competing compatible candidate).
But other participles are based on the present stem and do have their own exponents, e.g. the present participle in -ɜg:

\[
< [n], \Phi \{ (\uparrow \text{VFORM}) = \text{PTCP.PRS} \}, \_ > \nu
\]

\[
\begin{bmatrix}
\text{PHON.REP} & /3g/ \\
\text{DEP} & \text{LT} \\
\text{TYPE} & \text{NOMINAL} \\
\text{HOST} & \begin{bmatrix}
\text{IDENT} & \text{NIECE} \\
\text{TYPE} & \text{VERBAL} \\
\text{STEM} & \text{PRS}
\end{bmatrix}
\end{bmatrix}
\]

\( \nu \) in this configuration can also have a separate exponent – the impersonal suffix ɜ (to be discussed further on).
Past stem form of “two-stem” verb səw- / səd- ‘go’
Unlike v and n, Pst is always functionally empty. We need it as a separate node to handle 3-stem verbs with regular past stems in -əd.

Otherwise, the structure of intransitive past-stem forms mirrors that of present-stem forms, without sandwiching in nominal derivation.

The intransitive past stem of such verbs can also serve as the participle through Pac-Man spanning, because the rule \([nP \rightarrow vP n]\) can still apply.
Two-stem verbs: Pac-Man spanning of participle

\[
\uparrow\downarrow \quad nP \\
\uparrow\downarrow \quad vP \\
\uparrow\downarrow \quad v \\
\uparrow\downarrow \quad Pst \\
\uparrow\downarrow \quad v \atop @INTR \\
\uparrow\downarrow \quad \sqrt{PRED}=\text{‘go’} \\
\uparrow\downarrow \quad səd \\
\]
“Three-stem” verb

TenseP

↑=↓

vP

↑=↓

Tense

(↑vform)=cntrf
(↑pers)=3
(↑num)=sg

↑=↓

v

Pst

↑=↓

v

(↑pred)=‘sleep’

↑=↓

@intr

χʷəšš-əd-aid

< [Pst], Φ{}, _ > ∨

[PHON.REP /əd/]

[DEP LT]

(CLASS INTR)

(STEM PST)

(TYPE VERBAL)

[HOS HOST]

[IDENT NIECE]

[TYP TYPE VERBAL]

[CLAS CLASS INTR]

[STEM PRS]

Oleg Belyaev
Ossetic verb morphology
24 July 2023
30 / 35
Participle

- “Three-stem” verbs only have present stems and participles stored as VIs:

\[
< [\sqrt{v}, \text{Pst}, n], \Phi \left\{ \left( \uparrow \text{PRED} \right) = \text{‘sleep’} \right\}, - > \nu
\]

\[
\begin{bmatrix}
\text{PHON.REP} & /\chi^w\emptyset\text{t}/ \\
\text{CLASS} & \text{TR} \\
\text{STEM} & \text{PST} \\
\text{TYPE} & \text{VERBAL}
\end{bmatrix}
\]

- Therefore, Ossetic verbs have at most two forms stored in the lexicon, even if they appear to have three.
The idea that finite forms are based on participles may seem ad hoc. However, Ossetic does have another category that has exactly this structure: the Impersonal

- PTCP.PST + 3 + TM
- tense-mood marker is (in most forms) identical to the 3sg form of the verb ‘be’

That the Impersonal is based on the participle is demonstrated by the behaviour of “three-stem” verbs:

- \(\chi^w\text{əšt-3-wə} \) ‘they sleep’
- \(*\chi^w\text{əšš-əd-3-wə} \)

Fun fact: the verb ‘do’ has a dedicated impersonal stem čənd (cf. ptcp. and pst. ko(n)d), so this is not a case of periphrasis.

\[
< [v], \Phi\{(\uparrow \text{VOICE}) = \text{IMPERS}\}, _ > \xrightarrow{\nu}
\]

\[
\begin{pmatrix}
\text{PHON.REP} & /3/ \\
\text{DEP} & \text{LT} \\
\text{STEM} & \text{PST} \\
\text{TYPE} & \text{VERBAL} \\
\text{HOST} & \begin{pmatrix}
\text{IDENT} \\
\text{NIECE} \\
\text{TYPE} \\
\text{NOMINAL} \\
\text{STEM} \\
\text{PST}
\end{pmatrix}
\end{pmatrix}
\]
The impersonal

Impersonal of the verb xʷəšš- / xʷəšš-əd- / xʷəšt ‘sleep’

TenseP

↑=↓

vP

↑=↓

nP

(↑ VOICE) = IMPERS

↑=↓

v

↑=↓

vP

↑=↓

nP

(↑ VOICE) = IMPERS

↑=↓

nP

Tense

(↑ VFORM) = PRS
(↑ VOICE) = IMPERS

↑=↓

v

↑=↓

Pst

(↑ PRED) = ‘sleep’

↑=↓

√

@INTR

χʷəššt-3-wə
Conclusions

- In this talk, I made an attempt at an LrFG analysis of stem alternations in Ossetic.
- To some extent, morphomic phenomena in LrFG can be captured by the features stem and class, but it seems that Ossetic requires more.
- This means that VIs must be able to serve as exponents to empty f-structures.
  - I do not see any formal obstacles to such an analysis;
  - however, care must be taken to avoid overgeneration.
- The placement of morphology below V makes LrFG a more elaborate counterpart to the traditional “sublexical nodes”.
- Thus, if LrFG violates LI, then LI does not seem to be a substantial issue.
- What is more interesting are constraints on possible exponents, spans and their mapping – hopefully, to be investigated in the future.
What are the advantages of LrFG, though, compared to a purely lexicalist account (e.g. PFM+LFG)?

A very similar analysis is possible in PFM through ample use of the Elsewhere Principle.

However, this would involve a set of m-features (participle, stem, etc.) that have no automatic mapping to f-structure features, which creates a layer of redundancy.

On the other hand, the proposal that some c-structure nodes have no f-descriptions associated with them is rather bold.

After all, we do not find such “morphomic” elements in the syntax (e.g. heads that appear only to satisfy a c-structure rule).

It could be stipulated that such nodes only appear below $X^0$, which amounts to admitting that a boundary between morphology and syntax does exist.

Perhaps we can arrive at a boundary between morphology and syntax by treating them both together? This could be an interesting research programme.


