Ossetic verb morphology in LrFG

Oleg Belyaev

Lomonosov Moscow State University Institute of Linguistics RAS

23 July 2023

LFG Conference Rochester University

Outline

- 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":
 - kзn- / ko(n)d- 'do'
 - s3w- / səd- 'go'
 - nəmaj- / nəmad- 'count'
 - šuz- / səʁd- 'burn'
 - *3ft- / 3ftəd-* '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 persent 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 straighforward 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'.

Two stems

Functions of Ossetic stems

- Ossetic is less radical, but stems cannot be assigned any temporal value.
- Present stem:
 - temporal forms: present s3w-ən 'I go', future s3w-zən-3n 'I will go';
 - modal forms: subjunctive s3w-on 'I would go', optative s3w-in 'I would like to go', imperative su 'go!';
 - present participles: s3w-3g, s3w-ag, s3w-g3;
 - future participle *s3w-inag*;
 - infinitive s3w-ən:
 - destinative participle s3w-3n 'for walking'.
- Past stem:
 - past səd-tan 'l went';
 - counterfactual səd-ain 'I would have gone';
 - "past" participle (+ nominalization) səd.
- Each tense-mood paradigm has its own set of endings
 - $\blacksquare \Rightarrow$ 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.) \rightarrow *ba-kod-ton* 'I did' (pfv.)
 - $f = \delta t on$ 'l wrote / was writing' (ipfv.) $\rightarrow n = f \delta t on$ 'l 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:

■ $l \Rightarrow g k \Rightarrow n$ - 'cut' (ipfv.) $\rightarrow a - l \Rightarrow g k \Rightarrow n$ - '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.

ba-kod-t-ain pv-do-tr-cntrf.1sg

kзп-а do-sвjv.3sg

s3₩-∂ go-prs.3sg PRED'DO'ASPECTPFVVFORMCNTRFPRED'DO'ASPECTIPFVVFORMSBJVPRED'GO'ASPECTIPFVVFORMPRS

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): $nad-t-ain \rightarrow nat:ain$ 'I would have beaten',
 - as devoicing: $\check{z}ayd$ -t-on $\rightarrow \check{z}axton$ 'I said' (cf. ptcp. $\check{z}ayd$),
 - or as nothing: $f \rightarrow st t on \rightarrow f \rightarrow st on$ 'I wrote'.
- Labile "minimal pair": counterfactual intr. šəγd-aid 'it would have burnt', šəχt-aid (*səγd-t-aid) 's/he would have burnt smth.'
- Past tense forms use a separate set of endings: šəγd-i 'it burnt', šəχt-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*3*w* / *k*^w∂*d* 'cry': tr. *k*^w∂*t*:*a* 'cried';
- wom- / womd- 'vomit': tr. womt:a 'vomited';
- žyor- / žyord- 'run': tr. žyort:a 'ran', etc.
- Around 30 intransitive verbs with transitive endings
 - sometimes with close semantic counterparts: *wajən* 'run' (intr.)
- The converse also happens, although less frequently: məš- / məšəd-'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. bər- / pst. bər-əd / ptcp. bərd 'crawl';
- prs. kaf- / pst. kaf-əd / ptcp. kaft 'dance';
- prs. žar- / pst. žar-əd / ptcp. žard 'sing';
- prs. waš- / pst. waš-əd / ptcp. wašt 'cry' (of animal);
- prs. lsww-/ pst. lsww-əd / ptcp. lswd 'stand';
- prs. qaž- / pst. qaž-əd / ptcp. qašt 'play';
- prs. *qav-* / pst. *qav-əd* / ptcp. *qavd* 'plan';
- prs. x3s- / pst. x3s-əd / ptcp. x3št 'hold';
- **•** prs. $x^w \partial s \delta \delta'$ pst. $x^w \partial s \delta' \delta' \partial s \delta' \delta'$ ptcp. $x^w \partial \delta \delta' \delta'$ sleep'.
- All these verbs have two properties in common:
 - they are (morphologically) intransitive;
 - the past stem uses the "regular" suffix -*>d* 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 $x^w \partial s \delta - x^w \partial s \delta - x^w \partial s \delta + x^w \partial \delta + x^w \partial \delta + x^w \partial s \delta + x^w \partial$

	present		preterite		counterfactual	
	sg.	pl.	sg.	pl.	sg.	pl.
1	x ^w əšš-ən	х ^w əšš-зт	x ^w əšš-əd-t3n	x ^w əšš-əd-əšt3m	x ^w əšš-əd-ain	x ^w əšš-əd-aikkam
2	x ^w əšš-əš	x ^w əšš-ut	x ^w əšš-əd-t3	x ^w əšš-əd-əštut	x ^w əšš-əd-aiš	x ^w əšš-əd-aikkat
3	x ^w əšš-ə	x ^w əšš-əns	x ^w əšš-əd(-i(š))	x ^w əšš-əd-əštə	x ^w əšš-əd-aid	x ^w əšš-əd-aikkoj
	x 033 0		x 833 84(1(3))	X 833 84 8318	x 633 60 alu	

participle: x^wəšt

■ inflection of the transitive verb *šur- / šərd* 'chase'

	present		preterite		counterfactual	
	sg.	pl.	sg.	pl.	sg.	pl.
1	šur-ən	šur-зт	šərd-t-on	šərd-t-am	šərd-t-ain	šərd-t-aikkam
2	šur-əš	šur-ut	šərd-t-aj	šərd-t-at	šərd-t-aiš	šərd-t-aikkat
3	šur-ə	šur-əns	šərd-t-a	šərd-t-oj	šərd-t-aid	šərd-t-aikkoj
-	rticiple: š		Səru-t-a	səru-t-oj	Səru-t-alu	SƏTÜ-L-AIKK

Analysis Puzzles

- 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

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_RFG

- L_RFG 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 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 ν 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
 - $< [\sqrt{]}, \Phi\{(\uparrow \mathsf{PRED}) = \mathsf{`see'}\}, \{\mathbf{see} : (\uparrow \mathsf{OBJ})_{\sigma} \multimap (\uparrow \mathsf{SUBJ})_{\sigma} \multimap \uparrow_{\sigma}\} > \xrightarrow{\nu} see$
- 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 v":
 - $< [\sqrt{v}, v], \Phi\{(\uparrow \mathsf{PRED}) = \mathsf{know'}\}, \{\mathsf{know} : (\uparrow \mathsf{OBJ})_{\sigma} \multimap (\uparrow \mathsf{SUBJ})_{\sigma} \multimap \uparrow_{\sigma}\} > \stackrel{\nu}{\rightarrow} znaj$ -
- 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 redden, *to red. When it is not available, the verbal stem undergoes Pac-Man spanning: to orange, *to orangen.

More detailed v-structure

$\langle \dots, \dots \rangle \xrightarrow{\nu}$	PHON(OLOGICAL) REP(RESENTATION) P(ROSODIC)FRAME P(ROSODIC)LEVEL DEP(ENDENCE) CLASS TYPE	phonological realization & condition prosodic unit 1 2 {LT, RT} {inflectional classes} VERBAL NOMINAL ADJECTIVAL	
(,,, /	HOST	$\begin{bmatrix} \text{IDENT}(\text{ITY}) & \text{AUNT} & \text{ NIECE} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\$	

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 c[onstituent]structure tree and each leaf corresponds to one and only one c[onstituent]-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 X⁰ can be stipulated to be lower than above X⁰ (as with X⁰ 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 X⁰ 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:

 $N \rightarrow N_{stem} \; N_{num} \; N_{case}$

When affixes appear above X⁰, 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 $\sqrt{v} + v$ 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.

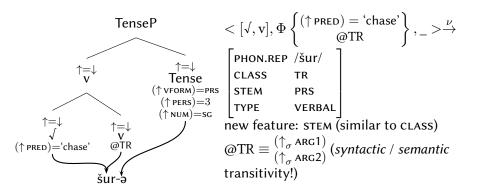
Preliminary remarks

- 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 $\uparrow = \downarrow$):

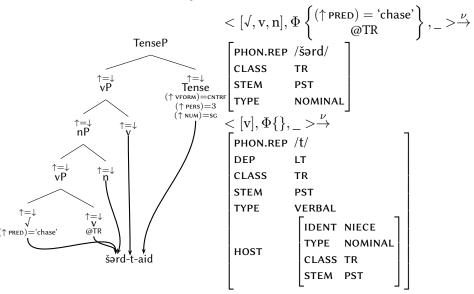
TenseP \rightarrow vP Tense (TAM endings) vP $\rightarrow \overline{v}$ Pst (past stem suffix) $\overline{v} \rightarrow \sqrt{v}$ (present stem, always a span) nP \rightarrow vP n (participles) vP \rightarrow nP v (from participle to verb stem)

- 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



Structure of transitive past-stem forms



• Counterfactual endings are agnostic with respect to transitivity:

$$< [\text{Tense}], \Phi \left\{ \begin{array}{l} (\uparrow \text{vform}) = \text{cntrf} \\ (\uparrow \text{pers}) = 3 \\ (\uparrow \text{num}) = \text{sg} \end{array} \right\}, _ > \stackrel{\nu}{\rightarrow} \left\{ \begin{array}{l} \text{Phon.rep /aid/} \\ \text{Dep } & \text{Lt} \\ \text{Stem } \text{PST} \\ \text{Type } \text{verbal} \\ \text{Host } \begin{bmatrix} \text{Ident niece} \\ \text{type } \text{verbal} \\ \text{stem } \text{pst} \end{bmatrix} \right]$$

"Functionally empty" heads

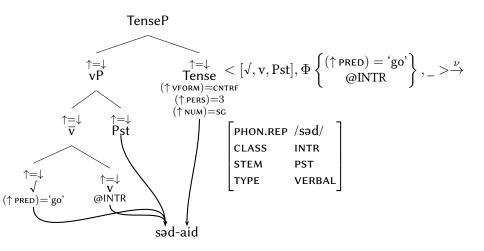
- The "nominalizer" n and the "verbalizer" v in finite transitive verbs are associated with no f-descriptions.
- This is not the case for n in general:
 - $(\uparrow v form) = ptcp.pst$
 - $(\uparrow vform) = ptcp.prs$
 - $(\uparrow vform) = ptcp.fut$
 - **...**
- The past participle has no independent 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 *-3g*:

$$<[n], \Phi\{(\uparrow \text{vform}) = \text{ptcp.prs}\}, _ > \stackrel{\nu}{\rightarrow} \left. \begin{array}{c} \text{phon.rep /3g/} \\ \text{dep } & \text{lt} \\ \text{type } & \text{nominal} \\ \text{host } & \begin{bmatrix} \text{ident niece} \\ \text{type } & \text{verbal} \\ \text{stem } & \text{prs} \end{bmatrix} \right]$$

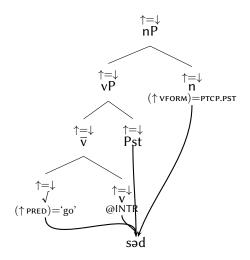
v in this configuration can also have a separate exponent – the impersonal suffix 3 (to be discussed further on).

Past stem form of "two-stem" verb s3w- / səd- 'go'



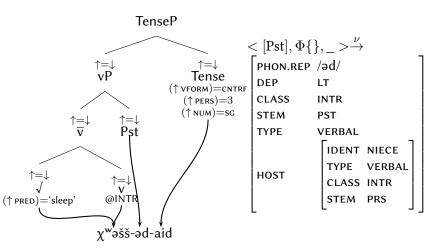
- Unlike v and n, Pst is always functionally empty. We need it as a separate node to handle 3-stem verbs with regular past stems in *-əd*.
- Otherwise, the structure of intransitive past-stem forms mirrors that of present-stem forms, without sandwiching in nominal derivation.
- The intransitive past stem of such verbs can also serve as the participle through Pac-Man spanning, because the rule [nP → vP n] can still apply.

Two-stem verbs: Pac-Man spanning of participle



Past stem: intransitive verbs

"Three-stem" verb



Participle

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

$$< [\sqrt{, v, Pst, n}], \Phi \left\{ \begin{array}{c} (\uparrow \text{ pred}) = `\text{sleep'} \\ @\text{INTR} \end{array} \right\}, _ > \stackrel{\nu}{\rightarrow} \left\{ \begin{array}{c} \text{PHON.REP} / \chi^w \hat{\text{o}st} / \\ \text{CLASS} & \text{TR} \\ \text{STEM} & \text{PST} \\ \text{TYPE} & \text{VERBAL} \end{array} \right\}$$

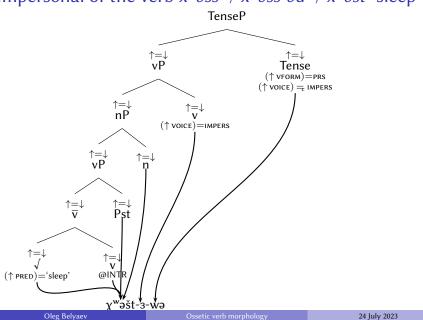
Therefore, Ossetic verbs have at most two forms stored in the lexicon, even if they appear to have three.

The impersonal

- 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
 - РТСР.РЅТ + з + ТМ
 - 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:
 - χ^wəšt-з-wə 'they sleep'
 - *χ^w 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.

$$< [v], \Phi\{(\uparrow \text{voice}) = \text{impers}\}, _> \stackrel{\nu}{\rightarrow} \left(\begin{array}{c} \text{Phon.rep } / \text{3} / \\ \text{dep} & \text{lt} \\ \text{stem} & \text{pst} \\ \text{type} & \text{verbal} \\ \text{host} & \begin{bmatrix} \text{ident niece} \\ \text{type} & \text{nominal} \\ \text{stem} & \text{pst} \\ \end{bmatrix} \right)$$

Impersonal of the verb $x^w \partial s \delta - x^w \partial s \delta - x^w \partial s \delta + x^w \partial \delta + x^w \partial \delta + x^w \partial s \delta + x^w \partial s \delta + x^w \partial \delta + x^w \partial \delta + x$



33/35

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 X⁰, 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 intereting research programme.

Abaev, Vasilij I. 1959. *Grammatičeskij očerk osetinskogo jazyka [A Grammatical Sketch of Ossetic]*. Ordžonikidze: Severo-Osetinskoe knižnoe izdatel'stvo.

- Anderson, Stephen R. 1969. "West Scandinavian vowel systems and the ordering of phonological rules." PhD diss., Massachusetts Institute of Technology.
- Asudeh, Ash, Tina Bögel, and Dan Siddiqi. 2022. "Modelling Exponents." Talk at Mo-MOT 2022, November 18, 2022.
- Belyaev, Oleg. 2021. "Paradigm structure influences syntactic behaviour: Ossetic case inflection." Chap. 16 in *Modular design of grammar*, edited by I Wayan Arka, Ash Asudeh, and Tracy Holloway King, 251–281. Oxford: Oxford University Press.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen, and Stephen Wechsler. 2016. *Lexical-functional syntax.* Second edition. Blackwell Textbooks in Linguistics 16. Oxford: Wiley Blackwell.

- Bresnan, Joan, and Samuel A. Mchombo. 1995. "The lexical integrity principle: evidence from Bantu." *Natural Language & Linguistic Theory* 13:181–254.
- Bybee, Joan, and Östen Dahl. 1989. "The creation of tense and aspect systems in the languages of the world." *Studies in Language* 13:51–103.
- Chistyakova, Daria. 2023. "Split (in)transitivity in Shughni." Master's thesis, Higher School of Economics, June 16, 2023.
- Everdell, Michael, Paul B. Melchin, Ash Asudeh, and Daniel Siddiqi. 2021.
 "Beyond c-structure and f-structure: On the argument-adjunct distinction in O'dam." In *Proceedings of the LFG21 Conference*, edited by Miriam Butt, Jamie Y. Findlay, and Ida Toivonen, 125–145. Stanford, CA: CSLI Publications.
- Lovestrand, Joseph, and John J. Lowe. 2017. "Minimal c-structure: Rethinking projection in phrase structure." In *Proceedings of the LFG'17 Conference*, edited by Miriam Butt and Tracy Holloway King, 285–305. Stanford, CA: CSLI Publications.

Lowe, John J. 2016. "English possessive 's: clitic AND affix." *Natural Language* & *Linguistic Theory* 34:157–195.

- Melchin, Paul B., Ash Asudeh, and Dan Siddiqi. 2020. "Ojibwe agreement in Lexical-Realizational Functional Grammar." In *Proceedings of the LFG'20 Conference*, edited by Miriam Butt and Ida Toivonen, 268–288. Stanford, CA: CSLI Publications.
- Tatevosov, Sergey G. 2019. "Aspekty slavjanskogo stilja" [Aspects of Slavic style]. *Voprosy jazykoznanija* 2:47–81.
- Toivonen, Ida. 2003. *Non-Projecting Words: A Case Study of Swedish Verbal Particles.* Dordrecht: Kluwer.
- Vydrin, Arseniy P. 2022. "Perexodnost' v osetinskom jazyke" [Transitivity in Ossetic]. *Acta Linguistica Petropolitana* 18 (3): 11–30.
- Wescoat, Michael T. 2002. "On lexical sharing." PhD diss., Stanford University.