Phrase structure and configurationality in Ossetic

Oleg Belyaev

Lomonosov Moscow State University

(Interdisciplinary School “Preservation of the World Cultural and Historical Heritage”)

Institute of Linguistics RAS

Pushkin State Russian Language Institute

Proceedings of the LFG’22 Conference

Miriam Butt, Jamie Y. Findlay and Ida Toivonen (Editors)

2022

CSLI Publications

http://csli-publications.stanford.edu/

Abstract

In this paper, I analyze Ossetic clause-level word order in terms of LFG. Ossetic word order is not constrained by the grammatical functions of elements, but by their formal features and information-structure functions. Specifically, interrogatives, negative pronouns, most subordinators and foci have to be positioned in the preverbal position. I argue that this pattern cannot be captured by the standard model of $\overline{\text{X}}$ theory because the limit to two bar levels does not provide enough "slots" for all these elements, and the notions of "complement", "adjunct" and "specifier" themselves become devoid of meaning. Instead, I propose that Ossetic has to be analyzed as having more than two bar levels; the notion of projection thus becomes similar to the notion of "slot" in template-based inflectional morphology. In effect, this solution suggests a view of clause structure as a language-specific phenomenon that develops through processes akin to grammaticalization in the domain of morphology.

1 Introduction

Ossetic is an Iranian language spoken in the Caucasus.[1] Like all neighbouring languages and most modern Iranian languages, the dominant word order in Ossetic is SOV. All other word order variants are attested as well, and there is no evidence in favour of a configurational structure associated with specific grammatical functions, like in English. However, Ossetic does severely constrain the linear position of a number of elements, namely preverbal negative pronouns, interrogatives and foci, and initial subordinators (C) and material that may precede these subordinators.

In this paper, I will argue that Ossetic word order is best treated in terms of discourse configurationality: preverbal material is organized in an $\overline{\text{X}}$-theoretic VP constituent, while the left periphery is positioned in a top-level CP constituent; a non-configurational S is sandwiched between these two strata. In terms of $\overline{\text{X}}$ theory, the number of "slots" that have distinct positioning rules is higher than that allowed by the traditional two-level $\overline{\text{X}}$ schema. Thus I argue that some languages, including Ossetic, may require more than two bar levels – something which is theoretically allowed by modern versions of $\overline{\text{X}}$ theory, such as Lowe and Lovestrand (2020), but has not, to my knowledge, been explicitly proposed in the literature. In a wider perspective, such a solution seems to eliminate any theoretical significance of the notions "specifier" and "complement", instead treating the $\overline{\text{X}}$ theory as a kind of syntactic "Item-and-Arrangement" model.

The article is organized as follows. In section 2, I provide a description of word order in Ossetic clauses, characterizing the constraints on the distribution of the

1. This research has been supported by the Russian Science Foundation, project no. 22-18-00528 "Clausal connectives in sentence and discourse: Semantics and grammaticalization paths". I am grateful to all native speakers who have provided consultations over the years starting from 2010, especially to Madina Darchieva. I would also like to thank the audience of LFG '22, especially Miriam Butt and Ash Asudeh, and two anonymous reviewers for helpful comments, and the editors for their hard work and patience. All errors are mine.

key elements which have fixed ordering: complex predicate components, negative pronouns, interrogatives and other preverbal foci, and subordinators. In 3, I interpret these data in terms of LFG, showing that the classical version of $\overline{\text{X}}$ theory is inadequate for capturing the full complexity of Ossetic word order. I conclude that for Ossetic, the restriction to only two bar levels, and consequently, the distinction between complements and specifiers, should be abandoned in favour of a model of endocentricity where bar levels are more similar to linear slots in template-based models of morphology.

2 Word order in Ossetic

2.1 Overview

While Ossetic is described as an SOV language, all theoretically possible orders are grammatical, as shown in (1).

(1) a. *žawər alan-ə fet:-a*
Zaur Alan-GEN see.PFV-PST.3SG
'Zaur saw Alan.'

b. *žawər fet:a alanə*
c. *alanə žawər fet:a*
d. *alanə fet:a žawər*
e. *fet:a žawər alanə*
f. *fet:a alanə žawər*

But Ossetic is far from being a nonconfigurational language in the way, for example, some Australian languages are (Austin and Bresnan 1996). It is more correctly described as discourse configurational with a grammaticalized preverbal area – in fact, the term discourse configurational was first introduced in É. Kiss (2004) for another language with a well-defined preverbal area, Hungarian. The basic facts on Ossetic word order can be found in Lyutikova and Tatevosov (2009), Erschler (2012), and Belyaev (2014).

In purely descriptive terms, the general organization of the Ossetic clause is summarized in (2). It can be divided into three fields: the verbal complex, i.e. the verb and the elements directly preceding it, which involves the most rigid linear constraints; the left periphery, i.e. the complementizer (Comp) and any preceding material; and the central area, i.e. elements between the left periphery and the verbal complex or to the right of the verbal complex – the least grammaticalized part of the clause, where constituents occur in relatively unconstrained order.

(2) $\text{PreC} - \text{C} - \underbrace{\text{PostC} - \overbrace{\text{Foc} - \text{Wh} - \text{Adv} - \text{Neg} - \text{NVC} - \text{V}}^{\text{verbal complex}} - \text{PostV}}_{\text{central area}}$

This ordering principle is complicated by second-position clitics, which appear after the first constituent in the clause regardless of its c-structure position or f-structure function. I will not discuss the syntax of clitics in this paper, but

their positioning will have a role to play in determining the c-structure configuration. An attempt at modelling Ossetic clitics in LFG can be found in Lowe and Belyaev (2015); for a general approach to clitics in LFG, see Bögel et al. (2010).

I will now demonstrate the ordering constraints captured in (2) one by one, starting from the bottom up.

2.2 The verbal complex

2.2.1 Complex predicates

In Ossetic, like in other Iranian languages, most lexical verbs are expressed as complex predicates, i.e. combinations of a nonverbal component (NVC) and a so-called light verb (most typically 'do', but also 'be', 'carry', etc.) that carries most morphosyntactic features.[2] The combination of NVC and the light verb is the most tightly bound grouping in the Ossetic clause. Preverbs normally attach to the NVC, not to the lexical verb: *ləg kɜn-ən* (cut do-INF) → ***a-****ləg kɜn-ən* (PV-cut do-INF). These complexes can only be split by second-position clitics; wh-words, which are also preverbal, cannot occur between the NVC and the light verb:

(3) *žawər* ‹✓ ***kɜ-imɜ*** › *ba-nəχaš* ‹* ***kɜ-imɜ*** › *kot:-a?*
Z. who-COM PV-speech do-PST.3SG
'**With whom** did Zaur speak?'

(4) *a-ləg =ɜj kot:-on*
PV-cut it.GEN do-PST.1SG
'I cut it off.'

Complex predicates without a preverb can behave both as a single unit or as the combination of a finite verb and a noun, i.e. the interrogative can be found to the left of NVC or between NVC and the light verb:

(5) *žawər* ‹✓***kɜ-imɜ***› *nəχaš* ‹✓***kɜ-imɜ***› *kot:-a?*
Zaur who-COM speech do-PST.3SG
'**With whom** was Zaur speaking?'

2.2.2 Negative pronouns

The next elements linearly closest to the verb are negative pronouns (6) or negation markers (7), which are in complementary distribution: standard Ossetic has no double negation.

(6) a. *žawər-ə* ***ni-či*** (*nɜ) *žon-ə*
Z.-GEN NEG-who NEG know-PRS.3SG
'No one knows Zaur.'

2. Complex predicates in Ossetic deserve a separate treatment. For LFG approaches to complex predicates, see e.g. Mohanan (1994) and Butt (1995). Like in all such cases, care must be taken to distinguish between idioms and "true" complex predicates. When speaking about light verbs and NVCs, I only mean the latter.

b. ***niči*** *žonə žawərə*

c. * ***niči*** *žawərə žonə*

d. * *žawərə žonə* ***niči***

(7) a. *žawər-ə* ***nɜ*** *žon-ən*
Z.-GEN NEG know-PRS.1SG
'I do **not** know Zaur.'

b. * ***nɜ*** *žawərə žonən*

c. * *žawərə žonən* ***nɜ***

When negative pronouns occur clause-initially, second-position clitics intervene between them and the verb, just as with complex predicates (8). But there are also two clitics which are second-position in affirmative sentences, but become preverbal in negative sentences: *=wal* 'more' and *=ma* 'yet' (9).

(8) *ni-sə* ***=jən*** *rat:-on*
NEG-what he.DAT give.PFV-PST.1SG
'I gave **him** nothing.'

(9) *žawər =ən* *ni-sə* ***=ma*** / ***=wal*** *rat:-a*
Z. he.DAT NEG-what yet more give.PFV-PST.3SG
'Zaur didn't give him anything **yet** / **more**.'

2.2.3 Preverbal foci

Interrogatives in Ossetic must occur preverbally, and narrow foci usually do as well, as seen from the question-answer pair in (10).

(10) a. *zul* [***či***]$_{\text{FOC}}$ *ba-lχɜt:-a?*
bread who PV-buy-PST.3SG
'**Who** bought bread?'

b. *zul* [***alan***]$_{\text{FOC}}$ *ba-lχɜt:-a*
bread alan PV-buy-PST.3SG
'**Alan** bought bread.'

While interrogatives are preverbal, they must precede negation and negative pronouns (and, by extension, NVCs of complex predicates), see (11).

(11) a. ***či*** ***ni-sə*** *žon-ə?*
who NEG-what know-PRS.3SG
'**Who** knows **nothing**?'

b. * ***ni-sə či*** *žon-ə?*

This also applies to preverbal subordinators (12), all of which are either synchronically or historically identical to interrogatives (see Erschler 2012).

(12) *kʷəd =zə š-fidar kɜn-ɜn iš iron ɜvžag,* ***kʷə ni-či*** *=jəl*
how it.IN PV-firm do-NMLZ is Ossetian language if NEG-who it.SUPER
nəχaš kɜn-a, wɜd?
speech do-SBJV.3SG then
'How is the Ossetic language to be strengthened **if no one** speaks it?'
(Ossetic National Corpus[3])

As for the relative positioning of interrogative and non-interrogative narrow foci, if the clause contains both a focused NP and an interrogative, the former must precede the latter or otherwise be postverbal:

(13) a. ***či*** *fet:-a* ***ɜrmɜšt žawər-ə****?*
who see.PFV-PST.3SG only Zaur-GEN
'**Who** saw **only Zaur**?'

b. ***ɜrmɜšt žawərə či*** *fet:a?*

c. * ***či ɜrmɜšt žawərə*** *fedta?*

Finally, mysteriously enough, adverbs in the comparative degree can also (optionally) intrude after the interrogative:

(14) *žawər-ɜj* ‹✓***rɜvz-dɜr***› *ni-či* ‹****rɜvz-dɜr***› *žʁor-ə.*
Zaur-ABL fast-COMPAR NEG-who run-PRS.3SG
'No one runs **faster** than Zaur.'

(15) *žawər-ɜj* ***či*** ***rɜvz-dɜr*** *žʁor-ə?*
Zaur-ABL who fast-COMPAR run-PRS.3SG
'**Who** runs **faster** than Zaur?'

This applies not only to adverbs, but also to NPs containing comparative adjectives:

(16) *səmɜ nɜ= qɜw-ə* ***či*** ***χʷəždɜr*** *lɜg u?*
I.wonder our village-IN who best man be.PRS.3SG
'I wonder who is the **best** man in our village?' (ONC)

The adverbs do not seem occupy the same position as NVC of complex predicates, because negative phrases and the verb cannot be broken up in this way:

(17) **žawər-ɜj ni-či* ***rɜvz-dɜr*** *žʁor-ə*
Z.-ABL NEG-who.NOM fast-ADD run-PRS.3SG
('No one runs faster than Zaur.')

2.2.4 Summary

The detailed structure of the verbal complex is summarized in (18).

(18) FOC – WH – ADV – NEG – NVC – V

Note that this is only a descriptive generalization; the c-structure of the verbal complex is discussed in section 3.3 below.

2.3 The central area

Outside the verbal complex, word order is much less constrained. The only constraint on material that precedes the verbal complex is that it cannot be focal (19b).

(19) a. [*žawər*]$_{\text{TOP}}$ [***alan-ɜn***]$_{\text{FOC}}$ *rat:-a* [***ɜχsa***]$_{\text{FOC}}$
Zaur Alan-DAT give.PFV-PST.3SG money
(What did Zaur give to whom?) 'Zaur gave Alan **money**.'

b. # [***alanɜn***]$_{\text{FOC}}$ [*žawər*]$_{\text{TOP}}$ *rat:a* [***ɜχsa***]$_{\text{FOC}}$

Postverbal material can be both topical and focal, in any order. Postverbal focus is illustrated in (13a). A nonfocal (topical or background, see Vallduví 1992) postverbal NP is illustrated in (20).

(20) *ɜmɜ sə fɜ-w-a, sə kɜn-a* [***lɜg***]$_{\text{TOP}}$*?..*
and what PV-be-SBJV.3SG what do-SBJV.3SG man
'And who should **the man** be, what shall (he) do?..' (ONC)

2.4 The left periphery

Most subordinators in Ossetic are preverbal (PV) and occupy the same position as interrogatives. But there are five subordinators which tend to be initial (nPV): *sɜmɜj* 'in order that', *salənmɜ* 'while', *kɜd* 'if', *səma* 'as if', *iwgɜr* 'if, since'. These can technically be located anywhere before the verb (up to the interrogative preverbal position), but in Belyaev (2014) I argued that they have two positioning variants: in C^0 or in the standard interrogative preverbal position. For example, the grammaticality of (21) degrades the farther the subordinator is positioned after the second position, but improves to normal when it is preverbal.

(21) ‹***sɜmɜj***› *žawər* ‹***sɜmɜj***› *rajšom* ‹$^{?}$***sɜmɜj***› *alan-ɜn* ‹$^{??}$***sɜmɜj***› *ɜχsa*
PURP Z. tomorrow A.-DAT money
‹***sɜmɜj***› *rat:-a* ‹****sɜmɜj***›
give-SBJV.3SG

'**so that** Zaur gives money to Alan tomorrow'

Corpus frequencies also confirm this generalization: in the vast majority of examples in the Ossetic National Corpus (even correcting for the length of the clause), nPV subordinators are initial, second-position, or preverbal; in fact, there is only one true example of a "third position" nPV subordinator in the sample of Belyaev (2014).

If it is true that these subordinators should be placed in a fixed position in C, whatever is to the left should be treated as located in the specifier or adjunct of CP. What purpose this dislocation serves in Osseic is unclear: it does not seem to be straightforwardly associated with topic status, for example. This question is outside the scope of this paper; what is important here is to establish the syntactic structure of the Ossetic clause, not the semantic or information-structure properties of its constituents as such.

3 Analysis

3.1 Overview

From the discussion above, it is clear that linear order in Ossetic does not encode grammatical functions. Therefore, following standard LFG logic, a traditional English-style configurational structure cannot be assumed; rather, Ossetic should be called a discourse configurational language in accordance with the typology of Snijders (2015, 144ff.). We are left with the following possibilities:

1. Fully flat S, with the order of elements defined via regular expressions.
2. A hierarchical clause structure, of which there can be the following variants
   (a) **VP – S – CP**: a three-level clause where VP is the verbal complex, CP is the left periphery, while S contains the rest of the material "sandwiched" between them. A possible problem with this approach is that the standard $\overline{\text{X}}$ model leaves us with too few dedicated positions for the preverbal elements whose position is thus constrained.
   (b) **VP – IP – S – CP**: same as above, but with an additional IP layer. This leaves enough $\overline{\text{X}}$-theoretic positions but requires motivation for lexical material to be present in I.
   (c) Unlimited $\overline{\text{V}}$ adjunction in the style of Japanese and Korean in Sells (1994, 1995). In this case, linear order should be constrained in some other way.
   (d) A cartographic approach in the style of Rizzi (1977) and subsequent work, with a host of separate projections (AspP, TenseP, AgrP, TopP, FocP, etc.). This is formally compatible with LFG architecture but is not typically applied, because projections are only stipulated when there is lexical material to fill them. L_RFG (Melchin, Asudeh, and Siddiqi 2020) introduces projections for inflectional features, but this is not completely equivalent to a full cartographic structure, and the specifier positions of these projections are not immediately identifiable with information-structure functions. Therefore, I will not consider cartography here, as such an analysis requires too many special assumptions to work in LFG.

In what follows I will demonstrate why a hierarchical anlaysis with VP, S and CP is to be preferred to other alternatives. A similar analysis of phrase structure to the one described herein has been used in Belyaev and Haug (2014), but that paper was focused on describing the syntax of correlatives and thus only assumed the structure, without much discussion or motivation.

3.2 Flat vs. hierarchical structure

As stated above, the simplest analysis is to assume a completely flat clause structure, but this would be inadequate, because there are at least two areas in the

Ossetic clause that display evidence of hierarchical organization: the verbal complex and the left periphery (i.e. the area adjacent to the initial subordinator).

3.2.1 Verbal complex

The verbal complex forms a prosodic word, or at least a prosodic phrase, with the verb (Abaev 1949). The exact acoustic characterization of the verbal complex has not been investigated, but the fact that immediately preverbal constituents are more tightly bound with the verb than other constituents is demonstrated by the fact that the initial *ɜ-* vowel of the verb is elided if the preceding element ends in a vowel (22). The same process occurs within noun phrases, which also form tight prosodic units (23), but does not occur across noun phrases (24).

(22) *sə (ɜ)r-səd-i?*
what PV-go-PST.3SG
'What happened?'

(23) *žawər-ə (ɜ)fšəmɜr*
Zaur-GEN brother
'Zaur's brother'

(24) *žawər-ə* *(ɜ)fšad *ba-jjɜft-a*
Z.-GEN army
'The army caught up with Zaur.'

Such prosodic facts cannot be taken as definite proof of constituency, because it is known that mismatches between prosody and syntax do occur. But, other things being equal, treating the verbal complex as a single constituent leads to a simpler mapping from c-structure to prosody, and should hence be preferred. However, Ossetic prosody is still in need of a separate, detailed study before definitive conclusions can be drawn.

But prosody is not the only motivation behind treating the left periphery and the VP as configurational. Another piece of evidence is coordination. One or more clauses can be coordinated under a single nPV subordinator; material to the left of the subordinator, if any, is shared across clauses (25).

(25) [$_{CP}$ *sɜlək:* ***sɜmɜj*** [$_{S}$ *amond-ǯən fɜndag-əl ba-ft-a*] *ɜmɜ* [$_{S}$
Ts. PURP happy-PROP way-SUPER PV-fall-SBJV.3SG and
iš-k^{w}ə aχ^{w}ər-mɜ fɜndag ššar-a]], ***wəj*** ***təχχɜj***
INDEF-when learning-ALL way find.PFV-SBJV.3SG that[GEN] for
'**So that** Tsælykk falls upon a happy path and one day finds a road to learning.' (ONC)

In contrast, all preverbal elements have to be repeated if more than one verb form is used (26). Scoping an interrogative over a coordinate phrase consisting of coordinated V + NP pairs is ungrammatical (27)–(28).

(26) *...ɜmɜ ɜž kɜsɜj žon-ən, məšajnag* [$_{VP}$ *či rat:-a*]
and I from.where know-PRS.1SG donation who give.PFV-PST.3SG
ɜmɜ [$_{VP}$ *či nɜ rat:-a*], *wəj.*
and who NEG give.PFV-PST.3SG that.DEM
'...and how do I know who gave the donation and who didn't?' (ONC)

(27) ***kɜm-ɜn** *rat:-a žawər činəg ɜmɜ ba-lɜvar kot:-a alan*
who-DAT give.PFV-PST.3SG Z. book and PV-gift do-PST.3SG A.
tetrɜd?
workbook
('**To whom** did Zaur give a book, and Alan present a workbook?')

(28) ***kɜj** *fet:-a alan ɜmɜ fe-q^{w}əšt-a soslan?*
who.GEN see.PFV-PST.3SG A. and PV-hear-PST.3SG S.
('**Who** did Alan see and Soslan hear?')

3.2.2 Left periphery

Were the clause a flat S, there would be no asymmetries between coordination patterns: either all clause subconstituents could be coordinated, or none. Therefore, we need at least a CP projection to explain why C can scope over coordinated constituents to its right, and a VP projection to capture the inseparability of the verbal complex. The material that is sandwiched between these projections can be assumed to be contained in a flat S, because the coordination of constituents containing such material is unconstrained: constituents to the left of the VP can be coordinated, sharing a right-adjacent constituent (29)–(30), or constituents to the right of the VP can be coordinated, sharing a left-adjacent constituent (31). I am aware of no evidence that could demonstrate that any of these constructions involve ellipsis in the form of Right Node Raising or a similar pattern. In a flat structure, this symmetricity of coordination can be accounted for by the finite-state mechanism of non-constituent coordination proposed in Maxwell and Manning (1996).

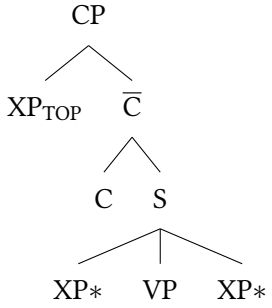
(29) *žawər-ə fɜnd-ə,* [$_{CP}$ *sɜmɜj* [*alan a-nəχaš kɜn-a*
Z.-GEN want-PRS.3SG PURP A. PV-speech do-SBJV.3SG
ažɜmɜt-imɜ$_i$] *ɜmɜ* [*šošlan š-sɜt:ɜ kɜn-a asɜmɜž-ɜn$_j$*
A.-COMIT and S. PV-ready do-SBJV.3SG A.-DAT
televizor] *šɜ=$_{i+j}$ xɜzar-ə*].
TV.set their house-IN
'Zaur wants **so that** [Alan speeks to Azæmæt$_i$] and [Soslan repairs for Atsæmæz$_j$ the TV] in their$_{i+j}$ house.'

(30) *žawər* [*alan-ɜn* *ba-lɜvar* *kot:-a*] *ɜmɜ* [*šošlan-ɜn* *a-wɜj*
Zaur Alan-DAT PV-present do-PST.3SG and Soslan-DAT PV-sale
kot:-a] *čing^{w}ə-tɜ*
do-PST.3SG book-PL
'Alan presented the books to Alan and sold (them) to Soslan.'

(31) *mɜn* *fɜnd-ə,* [$_{CP}$ *sɜmɜj* *wə-sə* *mašinɜ* [*š-sɜt:ɜ*
me.GEN want-PRS.3SG PURP that-ATTR car PV-prepared
kɜn-a *žawər*] *ɜmɜ* [*a-wɜj* *kɜn-a* *alan*]]
do-SBJV.3SG Zaur and PV-sale do-SBJV.3SG Alan
'I want [Zaur to repair] and [Alan to sell] this car.'

This leads us to adopt a c-structure for the Ossetic clause that roughly corresponds to (32).[4]

(32)



3.3 The structure of VP

The next question to consider is how to model the internal structure of the verbal complex, which we have established to be the VP. At least five distinct "slots" for preverbal material are required: (1) nonverbal components of complex predicates; (2) negative indefinites and negation; (3) comparative adverbs and NPs; (4) interrogative pronouns and complementizers; (5) focal non-wh NPs.

Just like at the top level of the clause, a flat VP structure is not an option, because coordination facts suggest that each class of elements is attached at its own level of projection, and coordination cannot violate the "levels" of the verbal complex:

(33) a. *či* [*ni-kɜj* *warž-ə*] *ɜmɜ* [*ni-kɜ-wəl*
who NEG-who.GEN love-PRS.3SG and NEG-who-SUPER
ɜwwɜnd-ə]?
trust-PRS.3SG
'Who loves no one and trusts in no one?'

4. An anonymous reviewer suggests using the label VC rather than VP for the verbal complex, in order to emphasize the difference between this style of "discourse configurational" VP and the "standard" VP that includes the verb and direct object. However, I prefer to keep the label VP as it captures the endocentric organization of this category.

b. *[*či* *ni-kɜj*] *ɜmɜ* [*kɜj* *ni-či*] *warž-ə?*
who NEG-who.GEN and WHO.GEN NEG-who love-PRS.3SG
('Who loves no one and is loved by no one?')

If both negative indefinites and interrogatives were VP complements, both would be non-constituent coordination, and there should be no difference, just as we have seen with S-level constituents above.

3.3.1 $\overline{\text{X}}$ schema

The only way to fit all these positions into the standard $\overline{\text{X}}$ schema for VP is to use adjunction to the largest extent and assume that NVCs of complex predicates are non-projecting words, as shown in the tree in (34).

(34) $[_{\text{VP}}\ \text{XP}_{\text{foc}}\ (1)\ [_{\text{VP}}\ \text{XP}_{\text{wh}}\ (2)\ [_{\overline{\text{V}}}\ \text{XP}_{\text{compar}}\ (3)\ [_{\overline{\text{V}}}\ \text{XP}_{\text{neg}}\ (4)\ [_{\text{V}^0}\ \widehat{\text{X}}_{\text{nvc}}\ (5)\ \text{V}^0\ ]\ ]\ ]\ ]\ ]$

This works for the data at hand, but turns $\overline{\text{X}}$ theory into nothing else but a system for labeling nodes in endocentric structures, and not a very convenient one at that. The distinction between complements, specifiers, and adjuncts, in particular, becomes completely blurred: most of the elements in question are iterable, and all except NVCs can be associated with any GF (argument or adjunct), but only $\text{XP}_{\text{COMPAR}}$ and NP_{FOC} are treated as adjuncts, for no other reason but to fit five elements into the $\overline{\text{X}}$ schema with three bar levels in total.

Futhermore, the tree in (34) assumes that NVCs of complex predicates are treated as non-projecting words, as in (35).

(35) $[_{\text{V}^0}[_{\widehat{\text{N}}}$*a-ləg*] $[_{\text{V}^0}$*kotː-a*]]
PV-cut do-PST.3SG

This is indeed plausible for most NVCs, which cannot be phrasal:

(36) a. ✓*zul-ə* *a-ləg*
bread-GEN PV-cutting
'the cutting of bread'

b. **zul-ə* *a-ləg* *kotː-on*
bread-GEN PV-cutting do-PST.1SG
('I cut the bread.', lit. 'I did the cutting of bread.')

However, some complex verbs, in certain contexts, can have a phrasal NVC, and it is located exactly where expected in the schema:

(37) ***sə*** [*dɜ=* *žɜrdɜ-mɜ*] *sɜw-ə?*
what thy heart-ALL go-PRS.3SG
'What do you like?' (lit. 'What goes to your heart?)

The most obvious solution would be to analyze (38) as an idiom, rather than a true complex predicate to be included in the verbal complex. However, this anlaysis would contradict the word order in (37), where the interrogative precedes the NP 'to your heart': this suggests that the latter is positioned rather low in the

verbal complex. Therefore, it seems that a full-fledged VP complement position is required to accommodate such NVCs, which should be treated as in (38).

(38) [$_{\overline{V}}$ [$_{NP_{nvc}}$ *dɜ= žɜrdɜ-mɜ*] [$_{V}$ *sɜw-ə*]]
thy heart-ALL go-PRS.3SG

If so,[5] the tree in (34) cannot be correct. Since bar level 1 is occupied by a complex predicate slot, we now have four **iterable** classes of elements (negative pronouns, comparative NPs, wh-words, other foci) that have to be distributed over **three** remaining positions: $\overline{V}$ adjunct, VP adjunct, and Spec-VP. Even if multiple specifiers are allowed (disregarding that SpecVP is a strange position for comparative adverbs), this does not fit into the schema: there is no available label for the uppermost node in (39), unless one resorts to parametrized rules (Dalrymple, Lowe, and Mycock 2019, 143–147; Frank and Zaenen 2002) for different types of adjuncts. But rule parametrization only serves to produce a family of different rules; in effect, it is not much different from the cartographic approach that stipulates constituents like TopP, FocP, etc., which is hardly acceptable in LFG for reasons mentioned above.

(39) [$_{?}$ XP_{foc} [$_{VP}$ XP_{wh} [$_{VP}$ XP_{compar} [$_{\overline{V}}$ XP_{neg} $\overline{V}$]]]]]

3.3.2 IP/VP split?

Another possibility is to introduce an IP layer into the clause structure, which gives us more than enough positions for all elements (40).

(40) [$_{IP}$ XP[foc] [$_{IP}$? [$_{\bar{I}}$ [$_{VP}$ XP[wh] [$_{VP}$ XP[compar] [$_{\overline{V}}$ XP[neg] $\overline{V}$]]] [INFL]]]]

But in LFG, constituents like IP cannot be taken for granted: it should first be demonstrated that the head node I can be filled by any lexical material. In fact, there is no evidence for IP in Ossetic. There are no periphrastic TAM paradigms. All periphrastic constructions that do exist, e.g. the periphrastic passive with *sɜw-* 'go' and the periphrastic resultative with *wə-* 'be' do not obey any fixed ordering. In (41), the passive auxiliary follows the lexical verb, while in (42) it is the other way around (although the former order seems to be preferred).

5. An anonymous reviewer observes that it is tricky to distinguish between complex predicates and idioms, and that constructions like (37) may turn out to belong to the latter class after all. In fact, the syntax of this construction does seem to be highly idiosyncratic: while examples like (37) are freely allowed, the interrogative may be immediately preverbal as well. Furthermore, it seems impossible to add a negative pronoun while retaining the ordering in (37): in this case both the interrogative and the negative have to appear before the finite verb. This strongly suggests that additional constraints, perhaps based on prosody, have a role to play in determining word order in such constructions. More research is needed on this topic, but the overall point is valid regardless of the analysis of complex predicates.

(41) *šk'ola-studi* «*bonvɜrnon*» ***arɜšt*** ***ɜr-səd*** 2001 *až-ə*
school-studio morning_star make.PTCP PV-go[PST.3SG] year-IN
cχinval-ə …
Tskhinval-IN
'The school-studio "Morning Star" was created in 2001 in Tskhinval …'
(ONC)

(42) 31 *dekabr-ɜj* *=nɜm* *abon-ə* *ong* *iw* *fədrakɜnd* *=dɜr* *ne*
december-ABL we.ALL today-GEN until one crime ADD NEG
r-səd ***arɜšt***
PV-go[PST.3SG] make.PTCP
'Not one crime **was committed** here (lit. "at us") from December 31st until today.' (ONC)

Therefore, adding I as a fixed position for the auxiliary verb, or assuming that the finite verb is sometimes found in I, does not lead to any helpful generalizations. Stipulation of IP in Ossetic is redundant.

3.3.3 Sells-style adjunction?

Another option is to adopt Sells' analysis of Japanese and Korean (Sells 1994, 1995). Sells assumes that $\overline{\text{V}}$ is the maximal projection in these languages. All verbal dependents are adjoined to $\overline{\text{V}}$; the Ossetic version of Sells' analysis would look like in (43).

(43) $[_{\overline{\text{V}}}$ *kɜj* $[_{\overline{\text{V}}}$ *ni-či* $[_{\overline{\text{V}}}\ [_{\text{V}}$ *fet:-a*]]]]
who.GEN NEG-who see.PFV-PST.3SG

Unlike the flat structure analysis, this correctly captures the coordination facts because binary branching is used. But the relative order of elements has to be captured by another mechanism, which could be f-precedence or rule parameterization. The former introduces additional complexity and duplicates the role of c-structure rules, which are supposed to simultaneously model linear order and dominance relations. The problems with parametrization were already discussed above.

3.4 Reinterpretation of $\overline{\text{X}}$ theory

What all these solutions (except Sells-adjunction, which is empirically problematic for Ossetic) have in common is that they use $\overline{\text{X}}$ theory in ways it was not originally meant to be. Perhaps, then, one should face the facts and reinterpret the role of $\overline{\text{X}}$-theoretic positions? The classical version of $\overline{\text{X}}$ theory uses three projection levels (unlike Jackendoff 1977, who used four) and, consequently, awards a central role to the distinction between complements and specifiers. This is also the case in the approach of Bresnan (2001), who associates specifiers with DFs and complements with non-discourse GFs. But we can see that the maximum bar level is a language-specific parameter: some languages, like Japanese and Korean, can apparently get away with just $\overline{\text{X}}$. Similarly, we can assume that some languages

require more than two levels if more than two endocentric positions are involved. In fact, this is already envisaged in the theory of Lowe and Lovestrand (2020), where the maximum level of projection is not viewed as universal.

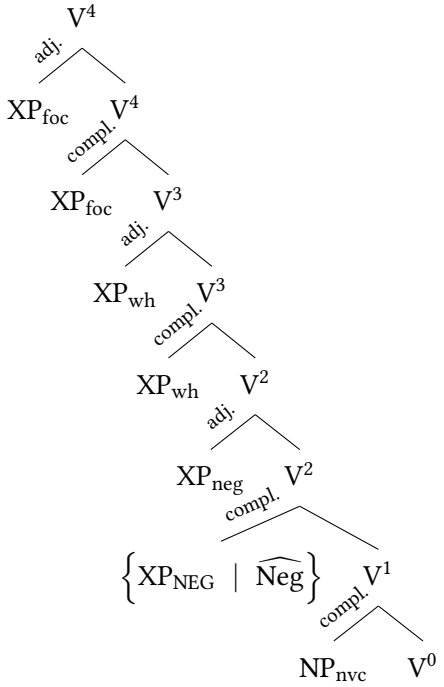
Under this interpretation, we need to distinguish between only two types of $\overline{\text{X}}$-theoretic rules: adjunction and complementation; adjunction does not increase the bar level, while complementation does. Both can apply at any bar level. In effect, this means that the role of $\overline{\text{X}}$ theory is the same as that of template morphology in languages with rich systematic affix ordering patterns. Consider the analysis of the Ashti Dargwa verb form in (44).

(44)

-4 nominal stem	**-3 preverb(s)**	**-2 negation**	**-1 gender**	**0 base stem**	**+1 causative**	**+2 … inflectional endings**
aq-	ha-	ta-	r-	ic:	-ah	-a-d-i
high	up	NEG	F	stand.PFV	CAUS	PRET-1-SG
'I did not make her stand up.'						

The role of the numbers in (44) is exactly the same as that of the projection levels in the "extended $\overline{\text{X}}$-theoretic" anlysis of the Ossetic VP in (45), viz. to capture the relative ordering and hierarchical embedding of dependents in endocentric structures (in syntax or in morphology).

(45)

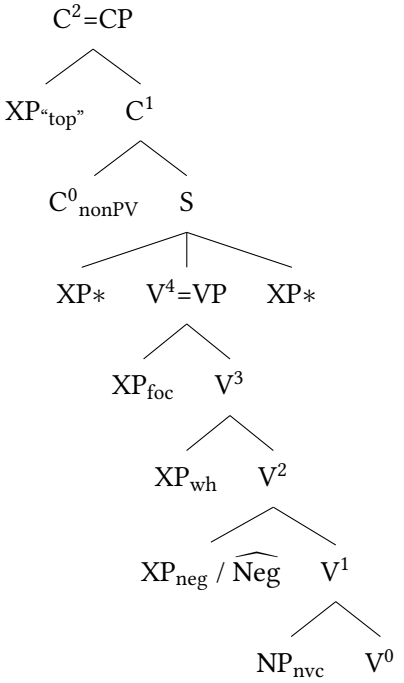


In this structure, each bar level corresponds to one type of preverbal element. This analysis is trivially compatible with Lowe and Lovestrand (2020). The special status accorded to final wh-words (Belyaev 2014) is captured by them being complement rather than adjuncts; this is extended to other classes of preverbal

XPs. Furthermore, I do not position comparative adverbs in this tree: they can adjoin to V^2 or have their own bar level, in which case the total number of bar levels in the Ossetic VP will increase to 5.

The new picture of the Ossetic clause as a whole that emerges from this approach is shown in (46).[6]

(46)



3.5 Rules

The formalization of this analysis is straightforward, although I cannot at present provide a rule for the lowest level (complex predicates), because no LFG analysis of Ossetic complex predicates has been elaborated yet. The overall system appears less complex than in Hindi/Urdu (Butt 1995): most complex predicates in Ossetic are lexicalized compound verbs, which should probably be represented by a single f-structure with a single argument structure, with the PRED value contributed by the non-verbal component. But Ossetic also has a productive system for forming deadjectival and denominal complex predicates, which is more difficult to model and deserves to be the object of a separate study.

The position of interrogative and negative pronouns can be modeled using constraining equations.[7] In (47), the V^2 rule introduces a defining equation for

6. Technically, the node S is endocentric, because in Ossetic it is always headed by V. It is therefore possible to treat S as an additional (flat) projection of V as well, as was proposed in Jackendoff (1977), who argued that S is V''' in English. But this would contradict established LFG practice, which treats all "flat" clausal projections with free GF assignment as S nodes.

7. I must admit that this analysis misses the fact that interrogatives and negatives can in fact occur outside of their canonical positions. For interrogatives, to my knowledge this only occurs in

negation, which is "checked" by the negative element itself. This guarantees that if the rule is used, this position must contain a negative element, and that negative elements themselves cannot appear outside of their dedicated positions. The machinery for interrogatives in (48) is the same. I use the f-structure function FOC to represent the position that is occupied by interrogatives and subordinator, as it is syntactically relevant in correlatives (Belyaev and Haug 2014). Other discourse functions should of course be represented at i-structure, as is current practice in LFG (King 1995; Choi 1999; Dalrymple and Nikolaeva 2011). For reasons of space, I omit the other rules, as they all involve either unification via $\uparrow = \downarrow$ (for $\overline{\text{X}}$ heads) or free assignment of grammatical functions via $(\uparrow \text{GF}) = \downarrow$.

(47) a. *ni-sə* D $(\uparrow \text{SPEC}) =_c \text{NEG}$
…

b. $\text{V}^2 \rightarrow$ XP $(\uparrow \text{GF})=\downarrow$ $(\downarrow \text{SPEC})=\text{NEG}$ V^1 $\uparrow=\downarrow$

(48) a. *sə* D $(\uparrow \text{SPEC}) =_c \text{WH}$
…

b. $\text{V}^3 \rightarrow$ XP $(\uparrow \text{GF})=\downarrow$ $(\uparrow \text{FOC})=\downarrow$ $(\downarrow \text{SPEC})=\text{WH}$ V^2 $\uparrow=\downarrow$

3.6 Similar proposals

Ossetic is not the only language that has a strictly organized preverbal area. In particular, the use of VP to describe an essentially discourse configurational structure (rather than the "standard" VP consisting of the verb and the direct object) has been suggested for Hindi/Urdu and Turkish in Butt and King (1996, 1997, 1999), on similar empirical evidence. Like Ossetic, Hindi/Urdu has preverbal focus, which Butt and King situate in the specifier of VP. The "verbal complex" in the narrow sense is represented as $\overline{\text{V}}$, which includes the verb together with certain auxiliaries and particles. The structure proposed by Butt and King for Hindi/Urdu is shown in (49).

(49) $[_{\text{IP}}\ [_{\text{IP}}\ \text{XP}_{\text{topic}}\ [_{\bar{\text{I}}}\ [_{\text{S}}\ \text{XP}_{\text{completive}}*\ [_{\text{VP}}\ \text{XP}_{\text{focus}}\ [_{\overline{\text{V}}}\ \text{V (V) (STAT) (AUX) }]\ ]\ ]\ \text{I}\ ]\ ]$
$\text{XP}_{\text{background}}*\ ]$

The difference between this structure and the structure I propose for Ossetic consists in two points. First, the structure of the VP in Ossetic must be much more elaborate than in Hindi/Urdu, because Ossetic has dedicated and distinct positions not only for foci, but also for interrogatives/subordinators and negative pronouns. Therefore, using only two levels with a flat $\overline{\text{V}}$ constituent is not sufficient. Second, the middle tier of the clause cannot be viewed as an IP constituent, because

the so-called multiple partitive construction (Haspelmath 1997, 177–179). Negative pronouns can appear postverbally under a kind of emphasis, especially if they are heavy, and in this case they trigger negative concord. I ignore these facts for now, as they deserve a separate treatment.

I find no evidence for a dedicated I position in Ossetic. Ossetic also allows more freedom in the assignment of discourse functions to postverbal material: it can be topical, focal, or backgrounded. For these reasons, and also due to lack of coordination asymmetries, I have opted to represent the "middle field" of the clause as a flat S layer.

Hungarian is another language of this type which has been extensively analyzed in LFG. Unlike Hindi/Urdu and like Ossetic, Hungarian has multiple dedicated preverbal positions for different classes of elements. It is thus not surprising that a very similar proposal for the c-structure of Hungarian predicates has been made in Börjars, Payne, and Chisarik (1999) and Payne and Chisarik (2000). Payne and Chisarik (2000, 204) propose the c-structure in (50) for Hungarian. The similarities with my analysis of Ossetic are in the use of more than two bar levels and in the special treatment of final foci, interrogatives and negative pronouns. Like in this proposal, the last elements of these classes are singled out as specifiers, while the rest are treated as adjuncts. The key difference is that Chisarik and Payne use Optimality Theory to model the relative order of some elements in the clause, while I propose that Ossetic word order can be described through phrase structure rules alone.

(50) $[_{V^3}$ QP $[_{V^3}$ QP $[_{V^2}$ $\begin{matrix}\text{FOC}\\ \text{INT}\\ \text{NEG}\end{matrix}$ $[_{V^2}$ $\begin{matrix}\text{FOC}\\ \text{INT}\\ \text{NEG}\end{matrix}$ $[_{V^1}$ $[_{V^0}$ NMR V] X(P)$*$]]]]]

Laczkó (2014) argues against the OT-LFG approach to Hungarian and proposes a different model for the VP, but his view of overall Hungarian clause structure in (51) is basically the same as my proposal for Ossetic: an S node sandwiched between CP and VP. Laczkò assumes a binary-branching S, but this is not essential: the same analysis could be used for Ossetic (with free left- or right-branching), given that coordination at this level is available in any order.

(51) $[_{CP}$ C $[_{S*}$ XP (T) $[_{S}$ XP (T) $[_{VP*}$ XP (Q) $[_{VP}$ XP (Sp) $[_{V'}$ V XP$*$]]]]]]

Thus, Hindi/Urdu and especially Hungarian share many similarities with Ossetic, but Ossetic is remarkable in having a larger number of dedicated positions within VP – something which cannot be dealt with in the standard $\overline{\text{X}}$ schema. My proposal is compatible with all thse approaches, because it uses $\overline{\text{X}}$ theory to represent any ordered hierarchical constituent structure.

4 Conclusions

In this paper, I have attempted to model Ossetic clause structure using standard LFG assumptions about c-structure. The problem with Ossetic word order is that the number of dedicated positions for different elements is so high that it requires either "going cartographic" or extending $\overline{\text{X}}$ theory to more than three projection levels, essentially eliminating "complement" and "specifier" as meaningful notions. Instead, what $\overline{\text{X}}$ theory models is which syntactic positions have grammaticalized in a given language. Just like lexicalist models of morphology do not

assume a fixed schema but accept that languages may grammaticalize different types of affixes in different orders, this view of phrase structure suggests that elements like VP and IP are not innate building blocks of grammar, but syntactic structures that grammaticalize differently in different languages. Ossetic, for instance, has developed discourse configurational VP and CP but has not developed any intermediate projections. Hungarian arrived at a similar structure, but with different elements included in the VP and in a different order. Classically configurational languages like English have taken a different route, grammaticalizing a VP and IP based on grammatical functions rather than discourse. This does not mean that these structures are completely random: again like in morphology, general laws of diachronic change mean that not all structures are equally likely to grammaticalize. This way of approaching configurationality seems to offer a viable alternative to approaches that assume a uniform hierarchy of projections (see, e.g., the treatment of Georgian preverbal focus in Skopeteas and Fanselow 2010 or of Hungarian topic and focus positions in É. Kiss 2008).

Many questions remain unanswered, both for Ossetic and in a more general sense. The status of S itself is unclear: if it is always headed by V, should it rather be treated as a projection of V? Or should S be used to represent a "flat" clause level regardless of its endocentricity? A full analysis of Ossetic clause structure also requires a separate, detailed account of complex predicates and negation, as well as the establishment of the information-structure status of elements preceding nPV subordinators. Periphrastic constructions should also be looked at in more detail for evidence of a hierarchical c-structure. Finally, what constituency tests except prosody and coordination can be used to determine clause structure in Ossetic? A weak point of the LFG approach to c-structure seems to be that very often, the number of analytical possibilities (flat vs. branching structure, adjunct vs. specifier) exceeds the number of criteria that can reliably distinguish between them.

References

Abaev, Vasilij I. 1949. *Osetinskij jazyk i fol'klor* [Ossetian language and folklore]. Vol. 1. Moscow and Leningrad: Izdatel'stvo Akademii nauk SSSR.

Austin, Peter, and Joan Bresnan. 1996. "Non-configurationality in Australian aboriginal languages." *Natural Language and Linguistic Theory* 14 (2): 215–268.

Belyaev, Oleg, and Dag Haug. 2014. "Pronominal coreference in Ossetic correlatives and the syntax-semantics interface." In *Proceedings of the LFG '14 Conference,* edited by Miriam Butt and Tracy Holloway King, 89–109. Stanford, CA: CSLI Publications.

Belyaev, Oleg I. 2014. "Korreljativnaja konstrukcija v osetinskom jazyke v tipologičeskom osveščenii" [Ossetic correlatives in typological perspective]. Candidate thesis, Lomonosov Moscow State University.

Bögel, Tina, Miriam Butt, Ronald M. Kaplan, Tracy Holloway King, and John T. Maxwell III. 2010. "Second position clitics and the prosody-syntax interface."

In *Proceedings of LFG'10 Conference,* edited by Miriam Butt and Tracy Holloway King. Stanford, CA: CSLI Publications.

Börjars, Kersti, John Payne, and Erika Chisarik. 1999. "On the justification for functional categories in LFG." In *Proceedings of the LFG '99 Conference,* edited by Miriam Butt and Tracy Holloway King. S: CSLI Publications.

Bresnan, Joan. 2001. *Lexical-Functional Syntax.* Blackwell Textbooks in Linguistics 16. Oxford: Blackwell.

Butt, Miriam. 1995. *The structure of complex predicates in Urdu.* Dissertations in Linguistics. Stanford, CA: CSLI Publications.

Butt, Miriam, and Tracy Holloway King. 1996. "Structural topic and focus without movement." In *Proceedings of the LFG '96 Conference.* Stanford, CA: CSLI Publications.

Butt, Miriam, and Tracy Holloway King. 1997. "Null Elements in Discourse Structure." Unpublished manuscript.

Butt, Miriam, and Tracy Holloway King. 1999. "Focus, Adjacency and Nonspecificity." In *Empirical Issues in Formal Syntax and Semantics 2: Selected Papers from the Colloque de Syntaxe et Semantique à Paris,* 75–98. The Hague: Thesus.

Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure.* Dissertations in Linguistics. Stanford, CA: CSLI Publications.

Dalrymple, Mary, John J. Lowe, and Louise Mycock. 2019. *The Oxford reference guide to Lexical Functional Grammar.* Oxford: Oxford University Press.

Dalrymple, Mary, and Irina Nikolaeva. 2011. *Objects and information structure.* Cambridge Studies in Linguistics 131. Cambridge: Cambridge University Press.

É. Kiss, Katalin. 2004. *The syntax of Hungarian.* Cambridge: Cambridge University Press.

É. Kiss, Katalin. 2008. "Topic and focus." *Acta Linguistica Hungarica* 55 (3–4): 287–296.

Erschler, David. 2012. "From preverbal focus to preverbal "left periphery": The Ossetic clause architecture in areal and diachronic perspective." *Lingua* 122 (6): 673–699.

Frank, Anette, and Annie Zaenen. 2002. "Tense in LFG." In *How we say WHEN it happens,* edited by Hans Kamp and Uwe Reyle. Berlin: Walter de Gruyter.

Haspelmath, Martin. 1997. *Indefinite pronouns.* Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.

Jackendoff, Ray. 1977. *X-bar syntax: a study of phrase structure.* Linguistic Inquiry Monographs 2. Cambridge, MA: MIT Press.

King, Tracy Holloway. 1995. *Configuring topic and focus in Russian.* Dissertations in Linguistics. Stanford, CA: CSLI Publications.

Laczkó, Tibor. 2014. "Essentials of an LFG analysis of Hungarian finite sentences." In *Proceedings of the LFG '14 Conference,* edited by Miriam Butt and Tracy Holloway King. Stanford, CA: CSLI Publications.

Lowe, John J., and Oleg Belyaev. 2015. "Clitic positioning in Ossetic." In *Proceedings of the LFG15 Conference,* edited by Miriam Butt and Tracy Holloway King, 229–249. Stanford, CA: CSLI Publications.

Lowe, John J., and Joseph Lovestrand. 2020. "Minimal phrase structure: a new formalized theory of phrase structure." *Journal of Language Modeling* 8 (1): 1–52.

Lyutikova, Ekaterina, and Sergei Tatevosov. 2009. "The clause internal left edge: Exploring the preverbal position in Ossetian." Handout, International Conference on Iranian Linguistics 3, University of Paris III: Sorbonne Nouvelle, Paris, September 11–13, 2009.

Maxwell, John T., III, and Christopher D. Manning. 1996. "A theory of non-constituent coordination based on finite-state rules." In *On-line Proceedings of the First LFG Conference,* edited by Miriam Butt and Tracy Holloway King, Rank Xerox, Grenoble, August 26–28, 1996.

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. S: CSLI Publications.

Mohanan, Tara. 1994. *Argument structure in Hindi.* Stanford, CA: CSLI Publications.

Payne, John, and Erika Chisarik. 2000. "Negation and focus in Hungarian: An Optimality Theory account." *Transactions of the Philological Society* 98:185–230.

Rizzi, Luigi. 1977. "The fine structure of the left periphery." In *Elements of Grammar,* edited by Liliane Haegeman, 281–337. International Handbooks of Linguistics. Dordrecht: Kluwer.

Sells, Peter. 1994. "Sub-phrasal syntax in Korean." *Language Research* 30 (2): 351–386.

Sells, Peter. 1995. "Korean and Japanese morphology from a lexical perspective." *Linguistic Inquiry* 26, no. 2 (Spring): 277–325.

Skopeteas, Stavros, and Gisbert Fanselow. 2010. "Focus in Georgian and the expression of contrast." *Lingua* 120 (6): 1370–1391.

Snijders, Liselotte. 2015. "The Nature of Configurationality in LFG." DPhil thesis, University of Oxford.

Vallduví, Enric. 1992. *The informational component.* New York and London: Garland.