ṚGVEDIC CLITICS AND 'PROSODIC MOVEMENT'

John J. Lowe
University of Oxford

Proceedings of the LFG11 Conference

Miriam Butt and Tracy Holloway King (Editors)

2011

CSLI Publications

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

Abstract

In this paper I develop an LFG account of second position clitic placement in Ṛgvedic Sanskrit. Clitic phenomena in this language are both more complicated and more ambiguous than (supposedly) in Serbian/Croatian/Bosnian, whose second position clitic data were recently treated by Bögel et al. (2010). I develop a formal treatment of clitic 'movement' which partly builds on Bögel et al.'s approach but which differs in certain fundamentals of formalism, maintaining a strict division between the syntactic and prosodic components of the grammar.

1 Introduction

There has been increasing interest in the interaction between prosody and syntax in recent years; recent work within LFG includes Butt and King (1998), Mycock (2006), and Bögel et al. (2009, 2010). One of the most problematic issues in the prosody-syntax interface is that of second position clitics, whose position in the clause, apparently determined at least partly by prosodic factors, cannot easily be accounted for under ordinary assumptions about syntactic constituency. An analysis of clitics within an LFG framework has recently been undertaken by Bögel et al. (2010); in this paper I consider the complex clitic data of Ṛgvedic Sanskrit and develop an alternative method of modelling the misalignment between syntax and prosody which this data reveals.

1.1 Ṛgvedic Clitics

The *Ṛgveda* is the earliest surviving text in the oldest Indo-Aryan language, Sanskrit; it is a collection of c. 1000 'hymns', metrical texts which originally had a ritual function and were composed c. 1500–1200 B.C. Ṛgvedic syntax is remarkably free: all possible orderings of V, S, and O are found, and word order is based on information structure (Viti, 2010); in fact word order patterns are very similar to those established for Ancient Greek by Dik (1995, 2007). Nevertheless there are clear syntactic rules and tendencies which provide evidence for structure within clauses; this is most apparent near the start of a clause, in particular in the position and ordering of clitic sequences.

According to Wackernagel's Law (Wackernagel, 1892), formulated specifically for ancient Indo-European languages like Ṛgvedic Sanskrit, unaccented elements

[†]I would like to thank Mary Dalrymple, Louise Mycock, Andreas Willi and Elizabeth Tucker for their invaluable comments and criticisms of earlier versions of this paper, also the audience at LFG11, in particular Dag Haug, for feedback and comments, and the audiences at the SE-LFG meeting at SOAS and at the Oxford Syntax Working Group who heard earlier versions of this paper. I gratefully acknowledge the financial support of the AHRC who have supported my doctoral study during which this work was undertaken, and in particular grants from my Faculty (Linguistics, Philology and Phonetics) and from the Lorne Thyssen Research Fund for Ancient World Topics at Wolfson College, which enabled me to attend the LFG11 conference.

occur in second position in the clause. However this is an over-simplified analysis of what is in fact a complex set of data. Exx. (1, 2) fit Wackernagel's pattern, but exx. (3, 4) apparently do not (clitics are underlined).

(1) *imáṃ ca no gavéṣaṇaṃ sātáye sīṣadho gaṇám*
this and our cow-seeking for_victory direct company
'And direct this our cow-seeking company to victory.' (6.56.5ab)

(2) *mó ṣú ṇaḥ soma mṛtyáve párā dāḥ*
not=and indeed us Soma to_death away give
'And do not hand us over to death, Soma.' (10.59.4a)

(3) *utá vā yó no marcáyād ánāgasaḥ*
also or who us would_harm innocent
'or also who would harm us innocent ones.' (2.23.7a)

(4) *divyā́ ā́po abhí yád enam ā́yan*
divine waters to when him came
'When the divine waters came upon him' (7.103.2a)

Note that not all enclitics are unaccented, and not all unaccented words are enclitic.[1] Several enclitics are accented, e.g. *hí* (in ex. 20 below), *sú* (in ex. 2), *nú*.[2] The major category of words in the *Ṛgveda* which is unaccented but not syntactically enclitic is finite verbs in main clauses, as *sīṣadho* in ex. (1) and *dāḥ* in ex. (2). Besides 'second position' clitics appearing to occur considerably later than second in the clause, there are other complications in the Ṛgvedic data which will be discussed below.

There have been two contrasting analyses of Ṛgvedic word order and in particular clitic placement in recent decades. A non-theoretical and largely descriptive 'Phonological Template' approach was developed by Hock (e.g. 1982, 1996), followed by Schäufele (1996). A transformational (GB) approach was developed by Mark Hale (e.g. 1987, 1996, 2007). The only attempt to analyse Ṛgvedic word order in a non-transformational theoretical framework is Schäufele's (1991) LFG-based asessment of RV syntax; but he did not adequately account for the positioning of clitics.

I will analyse the problems raised by the Ṛgvedic clitic data in §3, before developing my formal treatment of clitic misalignment in §4. Before that (§2) I will discuss the most recent approach to second position clitics in LFG, that of Bögel et al. (2010).

[1] Ṛgvedic accent was tonal, main word accent correlating with a high tone; the high tone is indicated with an acute in the transliteration.

[2] This is paralleled for example in Ancient Greek, where some second position clitics are accented and some are not, e.g. *ge*, *te* but *dé*, *oûn*.

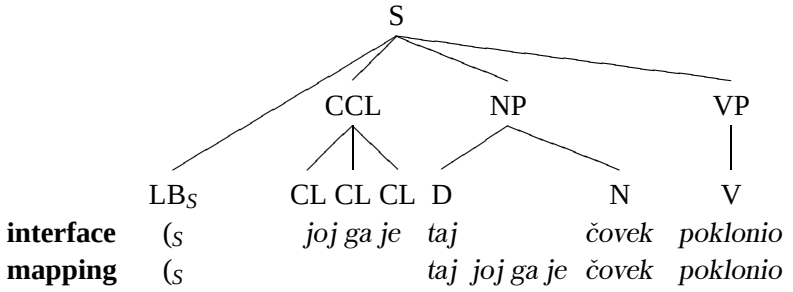
2 Clitic Sequences in LFG

The most recent treatment of second-position clitics in LFG is by Bögel et al. (2010), building on a new method of modelling the interaction between prosody and syntax presented in Bögel et al. (2009).[3]

Bögel et al. (2010) discuss problematic clitic phenomena in Serbian/Croatian/Bosnian (SCB), in which sequences of clitics (clitic clusters) appear within syntactic constituents. They explain this by assuming that clitic clusters can be generated in the c-structure at the left edge of a clause but cannot stand there in the prosody.

(5) *taj joj ga je čovek poklonio*
that her it AUX man presented
'That man presented her with it.' (Bögel et al., 2010, ex. 12a, p.112)

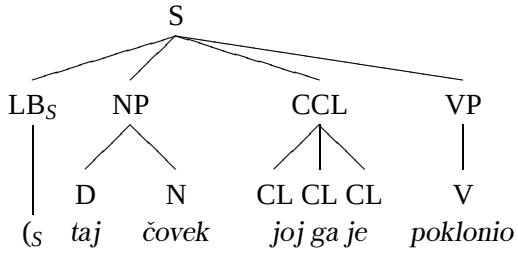
(6) C-Structure for ex. 5 (Bögel et al., 2010, ex. 26, p.118)



The clitic cluster *joj ga je* cannot stand in first position and so 'moves' to the right of the first accented element, the 'movement' equivalent to Halpern's (1995) Prosodic Inversion to account for second position clitics. In SCB the clitic sequence can also appear in syntactic second position, in which case it can be accounted for straightforwardly in the syntax.

(7) *taj čovek joj ga je poklonio*
that man her it AUX presented
'That man presented her with it.' (Bögel et al., 2010, ex.11a, p.112)

(8) C-Structure for ex. 7 (Bögel et al., 2010, ex. 39, p.123)



[3]Compare also Bögel et al. (2008) and Bögel (2010). The most important alternative approach to clitics is within the Lexical Sharing theory of Wescoat (2002, 2005, 2007, 2009), which I do not have space to discuss in detail; for a critical analysis of this theory in regard to clitic placement see Bögel (2010, p. 97–100).

In order to constrain the movement of clitic sequences, Bögel et al. (2010) introduce into the PS-rules and thereby the c-structure reference to prosodic boundaries, here the left edge of a clause (LB_S). They also introduce the CCL ('clausally-scoped clitic cluster') into the PS-rules such that it is generated next to the left edge of the clause, or as the second syntactic element in the clause. In order to control the 'phonological flip' they introduce an 'Interface Mapping Rule' which roughly states that a CCL clitic sequence can be placed one word further to the right in the prosodic output than it is in the syntax, so long as in the syntax it is directly to the right of a prosodic left boundary (and vice versa).

In my treatment I will adopt some of Bögel et al.'s fundamental assumptions; I will make use of the category CCL for some but not all clitics, and I share their intuition that the CCL can be generated at the start of a clause even if the clitics within it 'surface' in the prosody one word further to the right. However I will not adopt the more formal aspects of their approach. The projection of LB_S and RB_S in the c-structure significantly changes the nature of c-structure by allowing it to represent more than just the position and constituency of words. In their approach this is necessary, however, because the interface mapping rule needs to make reference to prosodic boundaries.

Moreover Bögel et al.'s interface mapping rule makes reference to clitic sequences appearing in the CCL such that phonological movement can only take place where the c-structure projects a CCL. In my view the 'movement' governed by the interface mapping rule is governed rather by the interaction of more general constraints on c-structure and p-structure formation which do not therefore require a special rule referring specifically to clitic groups appearing in a particular syntactic context. This also permits more complicated clitic data, such as that found in Ṛgvedic Sanskrit, to be accounted for without having to add to the formal architecture.

A third feature of the architecture formulated by Bögel et al. (2010) is the difference between prosodic and syntactic second position. Following Halpern (1995) they assume that Prosodic Inversion should be treated as a last resort: if the position of clitics can be accounted for in the syntax, then do so. SCB is usually taken as one of the best examples of a language in which the clitics *cannot* be positioned syntactically, because NPs such as *taj čovek* cannot otherwise be discontinuous.[4]

If, however, we have accepted the principle of phonological movement, it is questionable whether we can always treat it as a last resort. If it is a genuine linguistic possibility, then it may occur even when a fully syntactic account is equally valid. In particular, if the first syntactic constituent of a clause consists of a single prosodic word, then the difference between syntactic and prosodic second position

[4]The SCB data is more ambiguous than Bögel et al. (2010) and many earlier authors assume; see now Ćavar and Seiss (2011) and references therein. Nevertheless if such examples did not constitute a problem for syntactic analysis they would not have become so widely discussed; moreover the basic facts assumed for SCB by Bögel et al. (2010) are uncontroversially found in other languages such as Ancient Greek, Gothic, and possibly Warlpiri.

is neutralized.[5] In a language such as Ṛgvedic Sanskrit, where there is considerable freedom for discontinuous constituents, certain enclitics only ever occur after the first phonological word. Even where this appears to result in the enclitic 'interrupting' an intial constituent, it cannot be proven that we are not dealing with two discontinuous constituents with the clitic in second *syntactic* position, as in the following example.

(9) *imáṃ ca no gavéṣaṇaṃ sātáye sīṣadho gaṇám*
this and our cow-seeking for_victory direct company
'And direct this our cow-seeking company to victory.' (6.56.5ab)

Here it is tempting to take *imáṃ gavéṣaṇaṃ* 'this cow-seeking (thing)' as a single constituent, interrupted by two enclitics; however it is always possible in contexts like this to take the first word, here *imáṃ* 'this', as a separate constituent from the second non-clitic word. In other words there are contexts in which we cannot prove one way or another whether we are dealing with syntactic or prosodic second position. The assumption that syntax always takes priority over prosody has no absolute basis and cannot be used as a valid means of deciding a given case. If phonological movement is possible, this possibility should be independent of the syntax, i.e. it should be equally possible (within its own, phonological, constraints) in languages or contexts where our syntactic architecture can in principle permit a fully syntactic account of surface word order as it is in contexts where it cannot.

3 Clitics in Ṛgvedic Sanskrit

3.1 Clitic 'movement' in Ṛgvedic Sanskrit

Ṛgvedic Sanskrit, like many other languages, has a class of enclitic conjunctions which always follow the first non-enclitic (phonological) *word* of their domain, even where this entails interrupting what appears to be a syntactic constituent. As stated above, the considerable freedom for discontinuity of constituents means that it is hard to prove syntactic constituency and therefore clitic 'movement' in Ṛgvedic Sanskrit. But given the evidence for phonological movement in other languages, we do not necessarily *need* to prove syntactic constituency before we can assume movement.

3.2 Clitic conjunctions

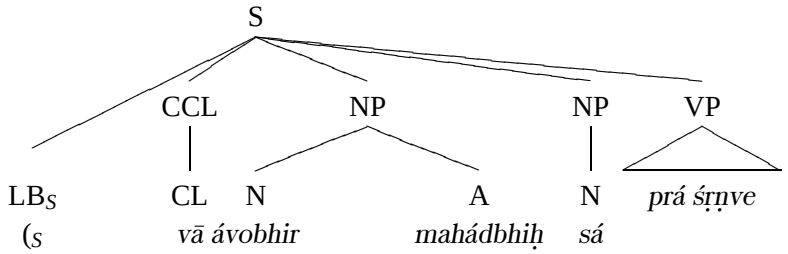
We will begin our survey of Ṛgvedic clitics by looking at clitic conjunctions. Descriptively, clitic conjunctions always appear in the output one word to the right of where they might be expected to be in the c-structure. According to Bögel et al. (2010, p.121, on Russian *li*) this involves a CCL, but one which can only appear in

[5] At least in terms of the output word order; there could potentially be prosodic differences between the two.

prosodic second position, not in syntactic second position. If their approach were applied to Ṛgvedic clitic conjunctions, the passage given in ex. (10) would have the c-structure given in ex. (11).

(10) *ávobhir* *vā* *mahádbhiḥ* *sá* *prá* *śṛṇve*
with_assistance or with_great this PRVB is_famed
'or through your great assistances this one is famed.' (4.41.2d)

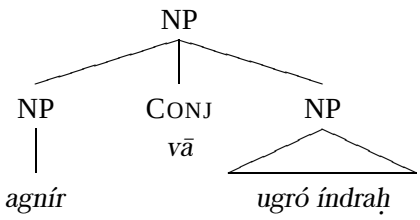
(11) C-Structure for RV 4.41.2d (ex. 10)



However in terms of linear order this is effectively the same position as for a 'normal', non-enclitic conjunction, i.e. directly preceding the first element of the conjoined S. In other words the CCL demanded by Bögel et al. (2010) to account for the position of enclitic conjunctions is in this case structurally in exactly the same position as the normal functional position of such conjunctions, CONJ.[6] It therefore makes more sense to treat the enclitic conjunctions as generated in CONJ rather than a CCL. The necessity of the CCL as a category will be discussed below, but we should follow the principle that as a non-standard X-bar category the CCL should be reserved only for clitics whose syntactic positioning cannot be treated under any ordinary syntactic category.

Enclitic conjunctions, of course, can also conjoin NPs, APs and PPs within clauses. We cannot suppose that a 'clausally-scoped clitic cluster' could appear within such constituents, rather the clitic conjunction appears in its regular position (CONJ). So for a simple NP conjunction such as *agnír ugró vā índraḥ* 'Agni or fierce Indra', the c-structure will be as follows.[7]

(12)



[6]This is not the case for Russian *li* which is the (subordinating) conjunction discussed by Bögel et al. (2010), but by extension would apply to conjunctions which are found in CONJ. Russian *li* can be generated in its syntactically expected position, C (cf. King, 1995, §10.2, p.232–238).

[7]This example from Hale (2007, p.205).

3.3 CCL Clitics

While clitic conjunctions always appear to the right of the phonological word corresponding to the terminal node which they directly precede in c-structure, other clitics, namely clitic pronouns and particles, have a variety of possible positions.

Firstly clitic pronouns and particles can occur cliticized to a governing word within a VP, NP or even PP; in these positions the clitics can be unproblematically generated under the appropriate syntactic node in the c-structure and no 'movement' need be assumed. This clitic position will therefore not be considered further. Such clitics can also appear near the start of a clause in a 'clitic cluster'. Descriptively, the start of any Ṛgvedic clause consists of a series of elements ordered according to a template of the following kind.

(13) (Conj) (XP) (XP) (Prvb) (Dem./Rel. Prons) (Pcls) (Prons)

All elements in the 'initial string' are in principle optional. The first possible element is a conjunction which cannot be preceded by any other element of the clause (assuming it is a clausal, not phrasal, conjunction); if the conjunction is enclitic it will follow the first word as discussed above. Then follow two positions which can be filled by any XP from the clause; these are usually considered to be topicalization/focus positions, and will be discussed in more detail below. Following this a preverb can occur, if it is not proclitic or enclitic on its verb; following this we find the regular position of the demonstrative and relative pronouns *sá-*, *syá-* and *yá-*. Then come enclitic sentence particles and finally enclitic pronouns. Since it is rare for more than one of the first five elements of the initial string to be filled, these enclitic words often appear in 'second position' in the clause. Following this initial string will be the rest of the sentence.

Previous analyses of the 'initial string' have taken the last two elements, the sentence particles and enclitic pronouns, as the only clitic categories (besides clitic conjunctions). However there is also evidence that preverbs and demonstrative/ relative pronouns should also be treated as clitic categories when appearing near the start of the clause (and not topicalized or focused). As this has not been previously recognized I will discuss this briefly before developing my analysis of CCL clitics.

3.3.1 Relative and Demonstrative Pronouns

Relative and demonstrative pronouns (including subordinating conjunctions which historically evolved from the former such as *yád* 'when, because') often occur in first position in the clause, in which case they can usually be interpreted as topical elements, i.e. descriptively in one of the XP slots in the initial string 'template'. However they can also appear following a different topicalized XP in which case they cannot be so analysed.

Evidence from sandhi phenomena suggests that in non-initial position demonstrative and relative pronouns were not infrequently enclitic. This, together with the fact that descriptively such pronouns directly precede the traditional 'clitic cluster',

suggests that in fact these pronouns should likewise be analysed within the clitic cluster.[8]

Sanskrit has internal and external sandhi rules. The latter govern the phonological interactions between independent phonological words which appear next to one another in the clause. Internal sandhi applies within phonological words, and for these purposes clitics do not count as independent phonological words, but part of the preceding or following word. This should give us a clear criterion for determining whether a given word is clitic or not in a given context; unfortunately internal and external sandhi differ in only a few details, so often we cannot be sure which we are dealing with. So for example the first segment of the relative pronoun *yá-* is not affected by sandhi and itself causes no distinct internal sandhi phenomena. The first segments of the demonstrative pronouns *sá-*/*tá-* and *syá-*/*tyá-*, on the other hand, are affected by internal sandhi rules. Specifically an *s* can be retroflexed according to the so-called 'ruki' rule, when one of the four segments *r*, *u*, *k*, *i* directly precedes it in the same phonological word. There are many examples of internal sandhi affecting the first segment of these demonstrative pronouns, which proves that, despite being accented in our texts and often occurring in initial position in the clause, these pronouns are, in these instances at least, enclitic on the preceding word. So we have *ṣá* for *sá* in ex. (14), *ṭáj* for *táj* in (15).

(14) *prá* *sú* *ṣá* *víbhyo* *maruto* *vír* *astu*
before PTC that from_birds Maruts bird let_be
'Let that bird be before (all other) birds, O Maruts.' (4.26.4a)

(15) *níṣ* *ṭáj* *jabhāra* *camasáṃ* *ná* *vṛkṣā́d* *bṛ́haspátir*
out that brought ladle like from_wood Bṛhaspati
'Bṛhaspati brought that out like a ladle from wood.' (10.68.8cd)

Although there is no sandhi evidence for the clitic status of the relative pronoun, there is evidence that related relative pronouns in other Indo-European languages were at least optionally enclitic. This is suggested by the positioning of relative pronouns in Old Irish (Watkins, 1963, p.29) and Ancient Greek (Fraser, 2001, p.141), the development of the definite adjective declension in Balto-Slavic possibly from a postposed relative construction (Vaillant, 1942), and the development of the ezafe construction from a relative pronoun in Iranian languages (Haider and Zwanziger, 1984; Haig, 2011).

We must therefore distinguish clitic and non-clitic variants of these words, the former generated under the appropriate XP nodes in cases of topicalization or focus, the latter generated in the 'clitic cluster' where not topicalized or focused.[9]

[8] Hale, among others, considers interrogative pronouns to be syntactically parallel to these relative and demonstrative pronouns but this is not in fact justified: interrogative pronouns can only ever be preceded by a single constituent while the others can be preceded by up to two constituents; by my analysis this means that interrogative pronouns must be topical elements filling the second optional XP of the clause (see above).

[9] The other enclitic pronouns also have non-clitic variants which are likewise used in contexts of

3.3.2 Preverbs

Further support for this position of relative and demonstrative pronouns comes from the positioning of preverbs. Preverbs are adverbial elements which can occur either near the front of the clause, or else directly preceding (proclitic on) the verb. Descriptively 'initial string' preverbs occur directly before the clitic cluster as formulated thus far, often first, or second following a 'topicalized' element (Hale, 1996, p.183–186).

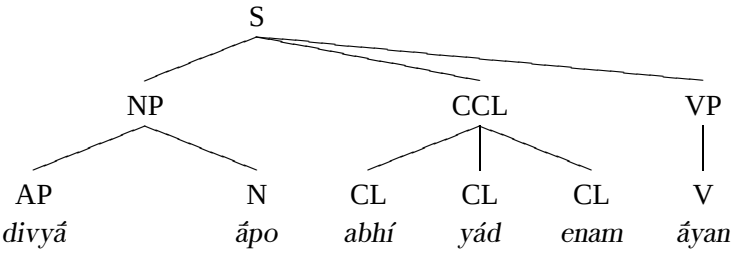
Preverbs at or near the start of a clause are usually treated as topicalized or focused elements. There are two competing justifications for this: either the preverb itself is focused, or the preverb serves to focus or topicalize the verb with which it is associated.

In some cases it is possible to treat the preverb as topicalized. However preverbs appear in their 'initial string' position in between 40–50% of their occurrences (depending on the particular preverb), and it is unlikely that directional adverbs would be topicalized or focused so frequently. Moreover statistical evidence shows that these initial string preverbs cannot (always at least) be serving to topicalize/focus the verb. For example there are c. 200 clauses with *yám* (accusative singular masculine relative pronoun) where the verb has no preverb - the verb precedes the relative pronoun (ie. it is topicalized/focused) in 3.5% of them; there are c. 70 clauses with a preverb - the preverb precedes the relative pronoun in 28.3% of them.[10]

We are left with preverbs in the initial string, directly preceding the clitic cluster, not topicalized or focused, words which in other contexts (when adjacent to the verb) are clearly (pro)clitics. If we assume that the CLL can host not just enclitics, but also proclitics, then the position of the preverbs can be easily explained.

(16) *divyā́ ā́po abhí yád enam ā́yan*
divine waters to when him came
'When the divine waters came upon him.' (7.103.2a)

(17) C-Structure for RV 7.103.2a (ex. 4=16)



topicalization and focus, but with these the difference is marked by the respective presence or lack of accent. Cf. also Selkirk (1995) on function words with accented and unaccented variants.

[10] I have taken statistics from clauses with relative pronouns since the topicalization/focus position before the pronoun is unambiguous, unlike in many other contexts.

This provides further evidence for the position of relative and demonstrative pronouns within the CCL, since in examples such as (4=16) the pronoun manifestly appears between CCL elements.

As proclitics, there is no restriction against preverbs occurring in clause-initial position: therefore it is possible for the elements of a CCL to remain in first position in a clause if the first element of the CCL is a proclitic. The CCL sequence of proclitic followed by one or more enclitics forms a single phonological word, within which internal sandhi rules apply regularly.[11]

3.4 The CCL

As argued in the previous section it is not always necessary to posit a CCL to host clitics which undergo prosodic inversion. However in other contexts this *is* justified. This is most clear, as in SCB, where more than one clitic of different lexical categories appear together in a particular position in the clause which cannot be justified on the basis of their lexical categorization. The same is true in Ṛgvedic Sanskrit, where clitic clusters in second position can include preverbs, sentence particles and enclitic pronouns.

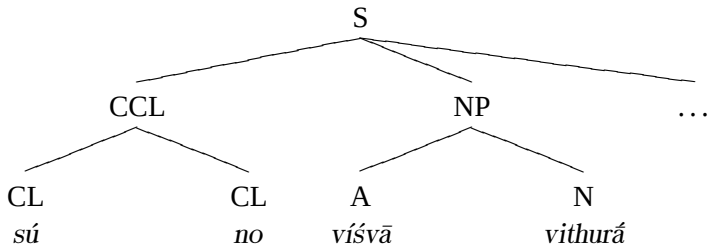
The syntactic treatment of clitic clusters is a problematic issue. Clitics which appear in clitic clusters are positioned on the basis of their clitic status rather than for any other syntactic reason. While clitic clusters often function as syntactic units (see e.g. Halpern, 1995, p.191–222 with references), there is no traditional XP category which can adequately dominate the varieties of clitics involved. The syntactic constituency of the CCL (and the fact that we have a CCL rather than a series of independent CLs) can only be based on the fact that the clitic cluster cannot be broken up by any other element of the clause. On the other hand clitic clusters could often be treated as single lexical items but do not seem to be formed in the lexicon according to normal morphological processes. Simpson and Withgott (1986) deal with clitic clusters by a process of 'template morphology' in the lexicon; the CCL utilized by Bögel et al. (2010) is an alternative, syntax-based approach. What both approaches share is the recognition that the syntactic constituency of clitic clusters cannot be accounted for by traditional X-bar theoretic rules. I will make use of the CCL here, but this approach could easily be adapted to alternative methods of treating clitic clusters.

(18) *víśvā* *sú* *no* *vithurā́* *pibdanā́* *vaso* *'mitrā́n*
all indeed for_us unstable firm good enemies
suṣáhān *kṛ́dhi*
easy_to_conquer make
'Indeed, make everything which is unstable firm for us, O good one, (and make) our enemies easy to conquer.' (6.46.6cd)

[11] Note that there is no constraint against a single phonological word having more than one accent: cf. lexical words such as *gnā́spáti-*, *bṛ́haspáti-* etc. The formation of single phonological words from proclitic plus enclitic is paralleled in Ancient Greek, with e.g. *eí te* > *eíte* etc.

When the CCL appears within another syntactic constituent, as in ex. (18), we can follow Bögel et al. (2010) and assume that in c-structure we in fact have a CCL at the start of the clause, which then undergoes 'movement' to after the first phonological word.

(19) C-Structure for ex. (18)



Sometimes a clitic pronoun or sentence particle (but not a conjunction) appears after the first *constituent*; again we can follow the principles proposed by Bögel et al. (2010) and generate the CCL in *syntactic* second position. The c-structure for the following example will then be exactly parallel to that given in ex. (8), but without the LB_S.

(20) *mahé* *kṣatrā́ya* *śávase* *hí* *jajñé*
to_great to_dominion to_might for was_born
'For he was born to great dominion (and) might.' (7.28.3c)

It is not necessary to utilize a PS-rule of the kind assumed by Bögel et al. (2010), involving optionality, to account for the variation between second prosodic and second syntactic position. Although the precise details for the Ṛgvedic Sanskrit are uncertain, I would rather assume a generic PS-rule such as the following, where the first optional XP in the clause is marked for focus or topicalization, meaning that the CCL appears in first position in the c-structure when nothing fills this first XP slot.[12]

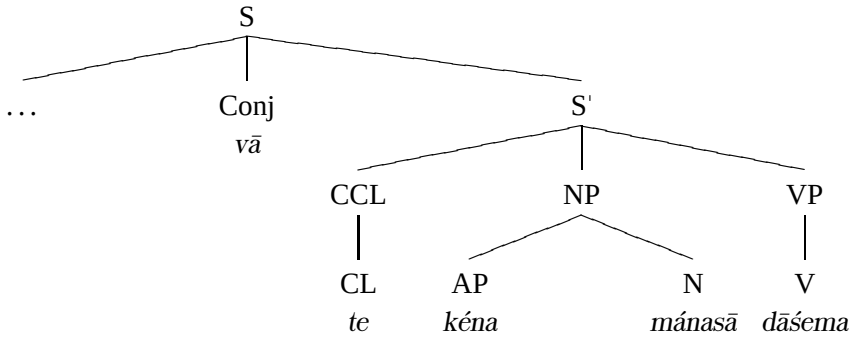
(21) S → (XP) (CCL) (XP)...

The following example shows both an enclitic conjunction, generated in CONJ, and an enclitic pronoun, generated in the CCL, which both undergo phonological movement to within the first constituent of the clause.

(22) *kéna* *vā* *te* *mánasā* *dāśema*
with_what or you with_attitude we_should_do_service
'or with what attitude should we do service to you?' (1.76.1d)

[12] In some passages it appears that the clitic cluster is positioned not in relation to the start of the clause but in relation to metrical boundaries such as line breaks and caesuras (see in particular Hock, 1996); this has been used to support a purely prosodic account of clitic placement in the *Ṛgveda*. However all such examples can be explained syntactically by assuming either dislocated topics or second syntactic position of the CCL, or both.

(23) C-Structure for RV 1.76.1d (ex. 22)



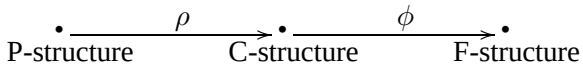
4 An OT analysis of ‘movement’

Previous approaches to clitic ‘movement’ involve addition to or alteration of the formal framework of LFG, whether this means adding further layers of structure as in Wescoat’s (2009) lexical sharing, or introducing prosodic brackets into c-structure as in Bögel et al.’s (2010) approach.

However it would be preferable to be able to deal with the positioning of clitics without arbitrarily adding to our theory.[13] We will see that the problem of clitic positioning can be resolved simply in an OT-LFG framework.

The positioning of clitics is a c-structure problem with a prosodic (p-structure) origin: the particular prosodic features of clitics which result in their appearance within constituents causes a problem when we try to account for their position in c-structure terms. In order to account for the position of clitics, then, we need a direct relation between c-structure and some kind of prosodic structure. A direct relation between c-structure and p-structure has been widely accepted since Butt and King (1998), found recently, for example, in Mycock’s (2010, p.292) schematic representation of the LFG projection architecture. A slightly different view is taken by Dalrymple and Mycock (2011) who assume only an indirect relation between c-structure and p-structure, mediated by the string of lexical items. But even such an indirect relation should be reconcilable with the OT-based account of ‘movement’ detailed here.

(24) Correspondence relations in the projection architecture (fragment)



If we assume a relation between p-structure and c-structure, we can annotate c-structure nodes with their respective relations to p-structure. This does not mean that every c-structure node directly corresponds to a p-structure node, though some

[13] Admittedly the CCL c-structure category is an addition to the traditional set of X-bar categories, but something special is required for the clitic cluster in any framework.

will. For the purposes of treating clitics, we require reference primarily to the level of the Prosodic Word, and for simplicity I will ignore other levels here.[14] Annotations in the c-structure can include the following.

(25)
- $\downarrow_\rho = \omega$: the corresponding element in the p-structure forms a phonological word.
- $\downarrow_\rho \subset \overrightarrow{\omega}$: the corresponding element in the p-structure forms a phonological word with the phonological word directly to its right.
- $*(_\omega \downarrow_\rho$: the corresponding element in the p-structure cannot stand at the start of a phonological word.

4.1 The OT constraint system

I assume that all structures in the grammar may be analysed according to their own set of OT constraints. In each case the GEN will be the rules of formulation of that structure, e.g. in the case of c-structure it will be the particular PS-rules of a language. All possible outputs from these structure-specific rules will form the input to the OT constraint systems. Assuming a direct mapping between c-structure and p-structure on the one hand, and c-structure and f-structure on the other, any OT approach to the c-structure will require coherence and faithfulness between the c-structure and the structures related to it.[15] Any given c-structure must in general correspond to coherent, possible, f- and p-structures, but more specifically it must always correspond to either a particular f-structure or a particular p-structure. That is, if we are parsing a given utterance, the c-structure must correspond to the particular, given, p-structure corresponding to that utterance, while if we are generating a sentence with a given meaning, the c-structure must correspond to the particular f-structure which corresponds to that meaning.

The clitic problem arises at the interface between the c-structure and the p-structure. Therefore if we have a given meaning (i.e. we are generating) and therefore f-structure, the c-structure will be generated on the basis of the f-structure without problem; the 'movement' of clitics will have to be accounted for in the inverse mapping from c- to p-structure (ρ^{-1}). If we are parsing, the p-structure will be given, and the 'movement' of clitics must be accounted for in the p- to c-structure mapping (ρ). I assume that the 'movement' is due to the interaction of competing prosodic and syntactic features which can best be represented in terms of OT tableaux.

For any given c-/p-structure pair, there will be one set of syntactic constraints governing the c-structures competing to represent a given p-structure (i.e. for ρ),

[14] I assume that, at least at this level of prosodic representation, the Strict Layer Hypothesis (Selkirk, 1984, p.26; Nespor and Vogel, 1986, p.7) is to be understood as a violable set of constraints (as per Selkirk, 1995). In particular, Phonological Words can be recursively formed, following e.g. Peperkamp (1996, 1997) and like Selkirk's (1995) 'affixal clitic' structure.

[15] Cf. Bresnan's (1996) 'C- to F-structure alignment' constraint.

and a different set of (phonological) constraints for the different p-structures competing to represent a given c-structure (i.e. for ρ^{-1}). I will not deal with the p-structure constraints here since they are a matter of phonology.

4.1.1 C-structure constraints

As stated above, the most basic constraints on c-structure involve faithfulness and coherence in relation to f- and p-structure. Faithfulness to f- and p-structures can be broken down into parts, e.g. FILL and PARSE, which penalize the addition and omission of input information, but for our purposes we can treat them as single constraints, which we will label F-ALIGN and P-ALIGN respectively, except that for explanatory clearness we will separate one particular sub-constraint of P-ALIGN and treat it separately.

(26)
- F-ALIGN: the c-structure is coherent with a possible/given f-structure.
- P-ALIGN: the c-structure is coherent with a possible/given p-structure.

The sub-constraint that we will treat separately is part of the requirement for coherence between c-structure and a given p-structure, namely that the integrity and order of phonological words be preserved in the mapping from p-structure to c-structure.

(27)
- ω-ALIGN: preserve the order and integrity of phonological words.

A similar constraint requires that the order of lexical items in the c-structure be the same as that in the corresponding p-string. I assume that s-forms are called into the c-structure by corresponding p-forms via the lexicon.[16]

(28)
- S-ORD: preserve the order of lexical items as found in the p-string.

All these constraints are high level constraints on competing c-structures which can be independently motivated and are not hypothesized purely to account for clitic 'movement'. The final constraint used here is likewise independently motivated. It is a constraint requiring economy of expression: see Morimoto (2001, p.171–172) for a discussion of contraints of this type. Morimoto's ECONOMY (derived from Bresnan's "Economy of Expression") constraint penalizes every XP and X^0 node in the c-structure; I follow his definition below.[17]

(29)
- ECONOMY: Economical structure is preferred (every XP and X^0 is penalized).

[16] The reference to linear adjacency required in this constraint could be formalized along the lines of Asudeh (2009), e.g. requiring that $\forall *.N(\rho(*)) = \rho(N(*))$ where ρ represents the mapping between p-structure and c-structure and N the next word according to linear sequence.

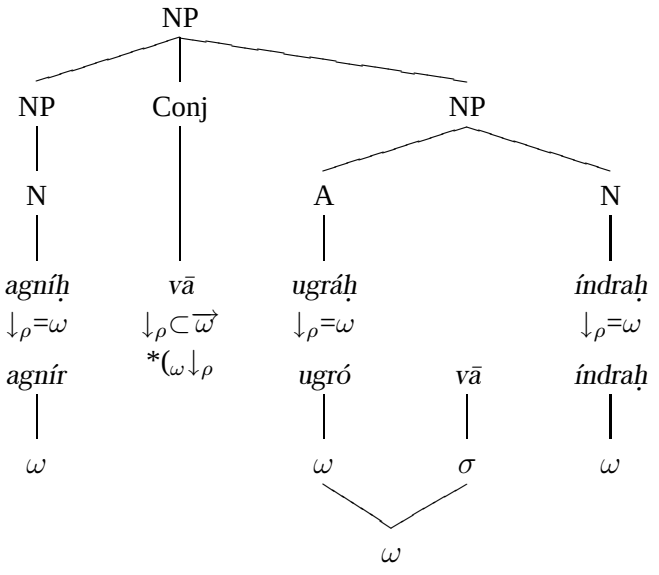
[17] Wescoat (2007, p.456) makes use of exactly the same constraint but labels it *PROJ.

Since all these constraints are independently motivated, this approach to clitic positioning does not add anything arbitrary to the grammar just to deal with a small group of problematic forms. Moreover, this approach is based on the given c-structure rules of a language. In principle this means that if the c-structure rules of a given language come to be more accurately understood such that 'movement' no longer need be assumed (this has happened with Warlpiri, for which the prosodic flip was once accepted) or indeed if an entirely new approach to c-structure is adopted in which clitic 'movement' is not a problem, the theoretical basis of this treatment is in no way invalidated or rendered superfluous. What it permits us to do is deal with the problematic positioning of clitics within the context of our current understanding of c-structure (in general and of given languages) without having to arbitrarily augment our theory.

4.2 Examples

We will take as the first example the NP *agnírugróvéndraḥ* 'Agni or fierce Indra' (ex. 12). To look at it first from the point of view of generation, and assuming a single rule for conjunction in Ṛgvedic Sanskrit, XP → XP (Conj) XP, we can draw the c-structure with annotations as below. Given the requirements of F-ALIGN it would be impossible to draw a tree in which *vā* followed *ugráḥ* which preserved the meaning 'Agni or fierce Indra', since if *ugráḥ* preceded *vā* it would have to be a modifier of Agni. The annotations on the lexical items in the c-structure permit the p-structure, drawn below, to be constructed. The annotations on the s-forms define their prosodic classification in the p-structure.

(30) *agnírugróvéndraḥ* (ex. 12)



(31) *agnírugróvéndraḥ* (ex. 12)

agníḥ ugráḥ vā́ índraḥ	F-AL.	P-AL.	ω-AL.	ECON.	S-ORD.
☞ a. *a. v. u. í.*				6	*
b. *a. u. v. í.*		*!		6	

When we build the c-structure on the basis of the p-structure derived from the output, we make use of the OT constraints as shown in the tableau.[18] Candidate (b) as given in the tableau preserves the order of lexical items as attested in the p-structure, while candidate (a) violates this constraint. Both candidates violate ECONOMY six times, so are not distinguished in this respect. However candidate (b), while preserving the order of lexical items, does not correspond to the given p-structure, since on the basis of the prosodic specifications of *vā́* it would, in that c-structure position, have to be forming (and not be at the start of) a phonological word with *índraḥ* rather than *ugráḥ*. Since the constraint P-ALIGN is ranked more highly than S-ORD, candidate (a) is the winner, even though this c-structure does not match the order of lexical items.

From this it should be clear that the constraint system is set up to require that the c-structure preserve the order of lexical items as found in the p-structure (and vice-versa) except in the very constrained context of clitics where there is no better option available.

In the following tableau, the ECONOMY constraint, being ordered before S-ORD, ensures that the preferred c-structure candidate is one in which neither clitic is treated in its s-string position even though it would be possible to leave *te* in its output position by assuming a discontinuous constituent.[19] The c- and p-structure trees are given in ex. (33).

(32) *kénavātemánasādāśema* (ex. 22)

kéna vā́ te mánasā́	F-AL.	P-AL.	ω-AL.	ECON.	S-ORD.
☞ a. *v. t. k. m.*				7	**
b. *v. k. t. m.*				8!	*
c. *v. k. m. t.*			*!	7	**

[18] Syntactic information has been omitted from the candidates in the tableau due to space restrictions; they should be read as follows:

a. [$_{NP}$ [$_{NP}$ [$_{N}$ *agníḥ*]] [$_{CNJ}$ *vā́*] [$_{NP}$ [$_{A}$ *ugráḥ*] [$_{N}$ *índraḥ*]]]

b. [$_{NP}$ [$_{NP}$ [$_{N}$ *agníḥ*] [$_{A}$ *ugráḥ*]] [$_{CNJ}$ *vā́*] [$_{NP}$ [$_{N}$ *índraḥ*]]].

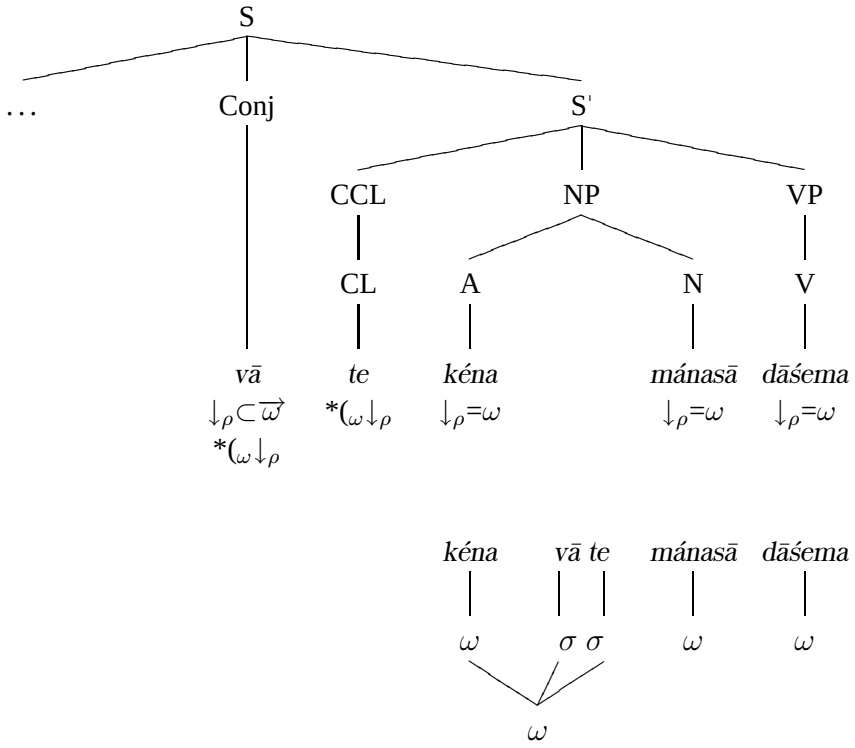
[19] Again the candidates have been abbreviated in the tableau; read as:

a. . . . [$_{CNJ}$ *vā́*] [$_{S}$ [$_{CCL}$ *te*] [$_{NP}$ *kéna mánasā*]. . .]

b. . . . [$_{CNJ}$ *vā́*] [$_{S}$ [$_{NP}$ *kéna*] [$_{CCL}$ *te*] [$_{NP}$ *mánasā*]. . .]

c. . . . [$_{CNJ}$ *vā́*] [$_{S}$ [$_{NP}$ *kéna mánasā*] [$_{CCL}$ *te*]. . .].

(33) *kénavātemánasādāśema* (ex. 22)



5 Conclusion

In this paper I have explored the clitic data of Ṛgvedic Sanskrit. I have adopted the CCL c-structure node from Bögel et al. (2010), but have shown that in the *Ṛgveda* not all clitics are generated in the CCL: clitic conjunctions are generated in their expected syntactic position. I have shown that contrary to traditional analyses preverbs and non-initial relative and demonstrative pronouns can be treated as clitics generated within the clitic cluster. My formal treatment of these clitic phenomena accounts for the apparent prosodic 'movement' of clitics in the interface between the c-structure and p-structure; OT constraints on the formation of both these structures govern the order of words in the respective structures and allow for the positional disjunction of clitics. The advantage of this analysis over that of Bögel et al. (2010) or Wescoat's lexical sharing hypothesis is that it works with minimal addition to the LFG architecture and unlike Bögel et al. (2010) it preserves the modularity of the grammar by keeping prosodic information out of the syntax.

References

Asudeh, Ash. 2009. Adjacency and Locality: a Constraint-Based Analysis of Complementizer-Adjacent Extraction. In Miriam Butt and Tracy Holloway King

(eds.), *Proceedings of the LFG09 Conference*, pages 106–126, Stanford, Ca.: CSLI Publications.

Bögel, Tina. 2010. Pashto (Endo-)Clitics in a Parallel Architecture. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG10 Conference*, pages 85–105, Stanford, Ca.: CSLI Publications.

Bögel, Tina, Butt, Miriam, Kaplan, Ronald M., King, Tracy Holloway and Maxwell III, John T. 2009. Prosodic Phonology in LFG: A New Proposal. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG09 Conference*, pages 146–166, Stanford, Ca.: CSLI Publications.

Bögel, Tina, Butt, Miriam, Kaplan, Ronald M., King, Tracy Holloway and Maxwell III, John T. 2010. Second Position and the Prosody-Syntax Interface. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG10 Conference*, pages 106–126, Stanford, Ca.: CSLI Publications.

Bögel, Tina, Butt, Miriam and Sulger, Sebastian. 2008. Urdu Ezafe and the Morphology-Syntax Interface. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, pages 129–149, Stanford, Ca.: CSLI Publications.

Bresnan, Joan. 1996. LFG in an OT Setting: Modelling Competition and Economy. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG96 Conference*, Stanford, Ca.: CSLI Publications.

Butt, Miriam and King, Tracy Holloway. 1998. Interfacing Phonology with LFG. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG98 Conference*, Stanford, Ca.: CSLI Publications.

Ćavar, Damir and Seiss, Melanie. 2011. Clitic Placement, Syntactic Discontinuity and Information Structure. This volume.

Dalrymple, Mary and Mycock, Louise. 2011. The Prosody-Semantics Interface. This volume.

Dik, Helma. 1995. *Word Order in Ancient Greek: A Pragmatic Account of Word Order Variation in Herodotus*. Amsterdam: J. C. Gieben.

Dik, Helma. 2007. *Word Order in Greek Tragic Dialogue*. Oxford: Oxford University Press.

Fraser, Bruce. 2001. The Clause Start in Ancient Greek: Focus and the Second Position. *Glotta* 77(3–4), 138–177.

Haider, Hubert and Zwanziger, Ronald. 1984. Relatively Attributive: The ‘ezāfe’-Construction from Old Iranian to Modern Persian. In Jacek Fisiak (ed.), *Historical Syntax*, pages 137–172, Berlin: Mouton de Gruyter.

Haig, Geoffrey. 2011. Linker, Relativizer, Nominalizer, Tense-particle: On the Ezafe in West Iranian. In Foong Ha Yap, Karen Grunow-Hårsta and Janick Wrona (eds.), *Nominalization in Asian Languages: Diachronic and Typological Perspectives*, pages 363–390, Amsterdam: John Benjamins.

Hale, Mark R. 1987. *Studies in the Comparative Syntax of the Oldest Indo-Iranian Languages*. Ph.D.thesis, Harvard University.

Hale, Mark R. 1996. Deriving Wackernagel's Law: Prosodic and Syntactic Factors Determining Clitic Placement in the Language of the Rigveda. In Aaron L. Halpern and Arnold M. Zwicky (eds.), *Approaching Second: Second Position Clitics and Related Phenomena*, pages 165–197, Stanford, California: CSLI Publications.

Hale, Mark R. 2007. *Historical Lingusitics: Theory and Method*. Oxford: Blackwell Publishing.

Halpern, Aaron. 1995. *On the Placement and Morphology of Clitics*. Stanford, Ca.: CSLI Publications.

Hock, Hans Henrich. 1982. Clitic Verbs in PIE or Discourse-Based Verb Fronting? *Studies in the Linguistic Sciences* 12(2), 1–38.

Hock, Hans Henrich. 1996. Who's on First? Toward a Prosodic Account of P2 Clitics. In Aaron L. Halpern and Arnold M. Zwicky (eds.), *Approaching Second: Second Position Clitics and Related Phenomena*, pages 199–270, Stanford, California: CSLI Publications.

King, Tracy Holloway. 1995. *Configuring Topic and Focus in Russian*. Stanford, Ca.: CSLI Publications.

Morimoto, Yukiko. 2001. Verb Raising and Phrase Structure Variation in OT. In Peter Sells (ed.), *Formal and Empirical Issues in Optimality Theoretic Syntax*, pages 129–196, Stanford, Ca.: CSLI Publications.

Mycock, Louise. 2006. *The Typology of Constituent Questions: A Lexical-Functional Grammar analysis of 'wh'-questions*. Ph.D.thesis, University of Manchester.

Mycock, Louise. 2010. Prominence in Hungarian: The Prosody-Syntax Connection. *Transactions of the Philological Society* 108(3), 265–297.

Nespor, Marina and Vogel, Irene. 1986. *Prosodic Phonology*. Dordrecht: Foris.

Peperkamp, Sharon. 1996. On the Prosodic Representation of Clitics. In Ursula Kleinhenz (ed.), *Interfaces in Phonology*, pages 102–127, Berlin: Akademie.

Peperkamp, Sharon. 1997. *Prosodic Words*. The Hague: Holland Academic Graphics.

Schäufele, Steven. 1991. *Free Word Order Syntax: The Challenge from Vedic Sanskrit to Contemporary Formal Syntactic Theory*. Ph.D.thesis, University of Illinois at Urbana-Champaign.

Schäufele, Steven. 1996. Now That We're All Here, Where Do We Sit? Phonological Ordering in the Vedic Clause-Initial String. In Aaron L. Halpern and Arnold M. Zwicky (eds.), *Approaching Second: Second Position Clitics and Related Phenomena*, pages 447–475, Stanford, California: CSLI Publications.

Selkirk, Elisabeth O. 1984. *Phonology and Syntax: the Relation between Sound and Structure*. Cambridge, Ma.: MIT Press.

Selkirk, Elisabeth O. 1995. The Prosodic Structure of Function Words. In Jill N. Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*, pages 439–469, Amherst, Ma.: GLSA.

Simpson, Jane and Withgott, M. Margaret. 1986. Pronominal Clitic Clusters and Templates. In Hagit Borer (ed.), *The Syntax of Pronominal Clitics* (volume 19 of *Syntax and Semantics*), pages 150–174, Orlando, Fl.: Academic Press.

Vaillant, André. 1942. L'article en vieux slave. *Revue des études slaves* 20, 5–12.

Viti, Carlotta. 2010. The Information Structure of OVS in Vedic. In Gisella Ferraresi and Rosemarie Lühr (eds.), *Diachronic Studies on Information Structure: Language Acquisition and Change*, pages 37–62, Berlin: De Gruyter.

Wackernagel, Jacob. 1892. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen* 1, 333–436.

Watkins, Calvert. 1963. Preliminaries to a Historical and Comparative Analysis of the Syntax of the Old Irish Verb. *Celtica* 6, 1–49.

Wescoat, Michael Thomas. 2002. On Lexical Sharing. Ph.D. Thesis, Stanford University.

Wescoat, Michael Thomas. 2005. English Nonsyllabic Auxiliary Contractions: An Analysis in LFG with Lexical Sharing. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG05 Conference*, pages 468–486, Stanford, Ca.: CSLI Publications.

Wescoat, Michael Thomas. 2007. Preposition-Determiner Contractions: An Analysis in Optimality-Theoretic Lexical-Functional Grammar with Lexical Sharing. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG07 Conference*, pages 439–459, Stanford, Ca.: CSLI Publications.

Wescoat, Michael Thomas. 2009. Udi Person Markers and Lexical Integrity. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG09 Conference*, pages 604–622, Stanford, Ca.: CSLI Publications.