PASHTO (ENDO-)CLITICS IN A PARALLEL ARCHITECTURE

Tina Bögel
Universität Konstanz

Proceedings of the LFG10 Conference

Miriam Butt and Tracy Holloway King (Editors)

2010

CSLI Publications

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

Abstract

This paper examines the phenomenon of Pashto (endo-)clitics, which are subject to both prosodic and syntactic contraints. These clitics challenge the view of prosody as being derivative from the syntax (e.g. Selkirk 1984) and the Principle of Lexical Integrity (Bresnan and Mchombo 1995) in that Pashto allows clitics to be inserted into the morphological word. However, these challenges can be resolved by assuming an architecture that views syntax and prosody as independent but interacting dimensions of grammar trying to align with each other as much as possible (see Bögel et al. (2009) for an approach within LFG). This paper presents data showing that it is the prosodic component that must account for the placement of the clitics *within* words, which leads to the conclusion that in cases of misalignment, the prosodic component takes precedence over the syntactic component, although this causes a violation of the Principle of Lexical Integrity.

1 Introduction

This paper examines the phenomenon of Pashto second position (endo)clitics as described by Tegey (1977).[1] Pashto is an Eastern Iranian language spoken in parts of Afghanistan and Pakistan by an estimated 40 million speakers. Clitics are quite common in this language; this paper mainly focuses on one group of second position (2P) clitics that have special properties which challenge the common understanding of the interaction of morphology, syntax and phonology. These clitics are subject to both syntactic and prosodic constraints and different approaches have been developed describing their placement in a clause. In general, Pashto clitics are placed following the first item of a sentence (the verb in (1a)). However, in the context of a stress alternation that accompanies a difference in aspect, this group of 2P clitics can appear as *endo*clitics — clitics that are placed *within* a word as in (1b):

(1) a. ṭakwαhə́ me
shake.IMPERF I
'I was shaking it.'

b. ṭák me wαhə
shake$_1$- I -shake$_2$.PERF
'I shook it.'

(Tegey 1977, 92)

Assuming that clitics are syntactic items in their own right, this phenomenon clearly poses a problem to the Principle of Lexical Integrity as stated in Bresnan (2001, 92), in that no syntactic item may intervene in a morphological word.

[†]I would like to thank Miriam Butt, Bariş Kabak, Tracy Holloway King, Astrid Krähenmann, Ghulam Raza and the audiences of the LFG conference and the MFG workshop for their help and their useful comments on the topic.

[1]For a general debate of 2P clitics see Halpern and Zwicky (1996) and references therein.

(2) **Lexical Integrity:**
Morphologically complete words are leaves of the c-structure tree and each leaf corresponds to one and only one c-structure node.

There have been different approaches to solving this problem. However, most of them are insufficient and do not fully account for the data. Thus, in this paper, I first describe the general 2P clitics and the prosodic and syntactic constraints that have to be considered (section 2). I then describe the phenomenon of endoclitics in Pashto and show that these are 2P clitics as well, following yet another prosodic constraint (section 3). In the case of Pashto endoclitics, prosody seems to be able to postlexically place a clitic after an accent-bearing element. The consequence is that prosody appears to be capable of overruling syntax and actually interacts with the morphological word. This phenomenon falls in with the assumption of a parallel architecture as introduced for LFG by Bögel et al. (2009), where prosody and syntax are assumed to be decoupled, but interacting modules of a grammar (following e.g. Zec and Inkelas (1990) and Lahiri and Plank (2009)).[2]

By assuming the syntactic and prosodic components to being parallel, clitics can be viewed as subject to prosodic and syntactic constraints simultaneously. This property is shared by clitics in other languages (e.g., Serbian/Croatian/Bosnian; see Franks and King (2000) for an overview), which seems to be natural, given that clitics have syntactic functions on the one hand and are prosodically deficient items on the other hand. By granting prosody an independent and strong position, I will show in section 5 how a morphological word can be interrupted by a clitic via the satisfaction of prosodic constraints, thus avoiding a violation of the Principle of Lexical Integrity.

2 Pashto Clitics

The group of 2P clitics discussed in this paper involves personal pronouns, modals and adverbials all of which are listed in Table 1:

Weak Pronoun	Num.&Pers.	**Modal**	Translation	**Adverbial**	Translation
me	1. Sg	ba	will, should	xo	really
de	2. Sg	de	should, let	no	then
ye	3. Sg				
am / mo	1. Pl				
am / mo	2. Pl				
ye	3. Pl				

Table 1: Pashto second position clitics and potential endoclitics

All of these clitics behave in the same way with regard to their position. If more than two clitics cooccur, they are placed in a fixed template, shown in (3).

[2]This architecture questions the architectural assumptions that view prosody as derivative of syntax, following the tradition of proposals made by Selkirk (1986) and Nespor and Vogel (1986); see also the criticism of general "syntactocentrism" by Jackendoff (2010).

(3)

1	2	3	4	5	6	7	8
xo	ba	am	am/mo	me	de	ye	no

(Roberts 1997, 5)

Note that in position 6, the clitic *de* can be only represented once even though this phonological shape can refer to two different clitics (see Table 1).

2.1 Syntactic Constraints

At first glance, these clitics seem to be common 2P clitics: they are placed after the first word of a sentence. Thus, in (4), the weak pronoun *ye* 'he' follows the noun *angur* 'grapes' while the modal clitic *ba* 'maybe' is placed after the adjective *nαroğa* 'sick' in (5).[3]

(4) angur ye rαwṛə
grapes he brought
'He brought grapes.' (Tegey 1977, 138)

(5) nαroğa ba wi
sick maybe is
'Maybe he is sick.' (Tegey 1977, 84)

However, the host of the clitic is not necessarily the first word of the sentence. Example (6a) shows that the element serving as a host for the clitic can be a syntactic constituent, in this case a coordinated noun phrase. This coordination may not be interrupted ((6b)):

(6) a. [xušαl aw patang]$_{NP}$ ba ye dər ta rαwṛi
Koshal and Patang will it you to bring
'Koshal and Patang will bring it to you.' (Tegey 1977, 84)

b. *[xušαl ba ye aw patang]$_{NP}$ dər ta rαwṛi

The same is true for postpositional phrases, where the clitic is not allowed in between the postposition and its argument:

(7) laylα na de αxistə (*laylα de na αxistə)
Layla from you buy
'You were buying it from Layla.' (Tegey, 1977, 114)

In constructions involving more than one sentence, the clitic may not appear outside of the clause in which it functionally originates, but instead is inserted after the first element therein. Thus, in (8), the two clitics are part of their individual clauses, occupying the second position respectively.

[3]Throughout the text, clitics are underlined.

(8) [tor me wəlidə] [magar [spin me wə nə lidə]]
Tor I saw but Spin I PERF not saw
'I saw Tor, but I didn't see Spin.' (Tegey 1977, 127)

Pashto provides numerous examples for the relatively regular syntactic placement of 2P clitics. However, since this paper is mainly concerned with the prosodic contraints and the interaction between prosody and syntax, the following sections will primarily focus on the influence of prosody on the placement of 2P clitics.

2.2 Prosodic Constraints

Up to this point, the constraints responsible for the positioning of the clitics in the above examples can be more or less explained syntactically. However, prosody plays a crucial part as well, as can be seen in (9) and (10), where the clitic is placed after the first item bearing lexical stress.

(9) rα ta te rα ṭolawə́l de
me for from_it here collect.IMPERF you
'You were collecting them for me from it (and bringing them) here.'
(Tegey 1977, 119)

(10) rα ta pe gα nḍə́ de
me for by_him sew.IMPERF you
'You were having him sew it for me.' (Tegey 1977, 119)

The elements preceding the verb belong to another group of Pashto clitics (Tegey's "Type II clitics"), which are usually placed in front of the verb. These clitics are all prosodically unstressed material, which forces the 2P clitic to appear after the first stressed element at the very right edge of the phrase even though Pashto is a fairly rigid verb-final language. The verb in the above constructions is the first element of the sentence bearing stress and hence the only proper host for the prosodically deficient clitics. Note furthermore that if the verb has constrastive or focus accent, the Type II clitics follow the verb. In these cases, the clitic in question is placed in between the verb and the Type II clitics ((11), cf. (10)) — all other positions are ungrammatical:

(11) gα nḍə́ de rα ta pe (*gα nḍə́ rα ta pe de)
'You were having him *sew* it for me.' (Hock 1996, 235)

Based on these examples one could argue that it is simply the head of the clause that the clitic attaches to. The following examples contradict this hypothesis in that the clitic is clearly attached to the stressed element, even though this element is *not* the head of the clause in (12a) (in contrast to (12b)):

(12) a. rα sará de wi
me with let be
'Let it be with me.'

b. rα sara wí de
me with be let
'Let it *be* with me.' (Tegey 1977, 121)

The hypothesis that the clitic attaches to the first accent-bearing element of the sentence is also confirmed when it comes to endoclitics. Although endoclitics are most common in simple verb-clitic constructions, they also appear in expressions where every element of the sentence apart from the verb is unstressed. This leads to an alternate version of example (9):

(13) rα ta te rα [ṭól de kṛəl]$_V$
me for from_it here collect$_1$ - you -collect$_2$.PERF
'You collected them for me from it (and brought them) here.'
(Tegey 1977, 119)

Here, the clitic is inserted *into* the verb following the part of the verb that bears the main accent and thus reacting to an verb-internal stress shift that comes along with a change in aspect, the main environment for the phenomenon of endoclisis as described in the following section.

3 Pashto Endoclitics

Like South-Asian languages in general, Pashto is an argument-dropping language (e.g. Butt (2007) and references therein). Sentences can therefore consist of only a verb and a clitic. The endoclitics mainly appear in these short sentences in the context of a stress alternation that accompanies a difference in aspect as in example (1), repeated here for convenience:

(14) ṭák me wαhe
shook$_1$ - I -shook$_2$.PERF
'I shook it.' (Tegey 1977, 92)

In Pashto, the *perfective* aspect of the verb is accompanied by a verb-internal stress shift placing the main stress on the first foot of the verb, while the verb in the *imperfective* aspect carries the main stress on the last foot of the verb. With regard to the stress shift, Pashto verbs fall roughly into three classes, depending on their word-internal structure. Since these structures are essential to the correct placement of the clitics, it is necessary to analyse them more closely in order to find the appropriate (prosodic or syntactic) unit on which the clitics depend. Thus, the different verb classes, their internal characteristics and their behavior concerning the placement of clitics will be introduced below.

3.1 Monomorphemic Class I Verbs

Class I verbs are monomorphemic. In the imperfective, these verbs bear stress on the last foot; the clitic is placed after the verb ((15a)). In the perfective aspect however, class I verbs take on a perfective prefix *wə-* which receives the main stress. In this case, the clitic occurs after the prefix and in front of the stem ((15b)):

(15) a. **imperfective**
təxnawə́la me
tickle I
'I was tickling (her).'

b. **perfective**
wə́ me təxnawəla (*wə́təxnawəla me)
PERF I tickle
'I tickled (her).' (Tegey 1977, 86)

3.2 Bimorphemic Verbs

In contrast to class I verbs, class II and III form the perfective by means of a stress shift from the last to the first foot of the verb without adding a perfective prefix. The verbs of both classes are bimorphemic. Class II verbs consist of a derivational prefix and a root. In the imperfective aspect, the stress is on the second foot of the verb — the clitic is placed after this ((16a)). The perfective is formed via a stress shift from the last to the first foot of the verb. The clitic is then placed after this first foot as in example (16b), i.e. after the derivational prefix.

(16) a. **imperfective**
ṭelwɑhə́ me
push I
'I pushed (it).'

b. **perfective**
ṭél me wɑhə (*télwɑhə me)
PREF I push
'I was pushing (it).' (Tegey 1977, 92)

Class III verbs are complex predicates consisting of a adjective, adverb or noun and a light verb and form the largest group of verbs in Pashto. Their behavior with respect to clitics is the same as with the class II verbs in that there is a verb-internal stress shift that goes along with a change in aspect, and that the clitic will be positioned after the first foot in the perfective ((17)) and after the whole verb in the imperfective.

(17) **perfective**
póx me kə
cook I do
'I cooked (it).' (Tegey 1977, 98)

With class III verbs, one can easily identify the single elements of the word because they are complex predicates. Thus, an analysis in favor of treating all three elements as postlexically independent items seems likely.

With class II verbs on the other hand, the separation of the elements is not as clear-cut, but one could argue that the derivational prefix might itself be a 'lexical word' (Anderson 2005), e.g. a clitic. Assuming that clitics are postlexical elements that occupy separate syntactic nodes, the class II verb in (16b) would thus lead to a c-structure representation similar to Figure 1:

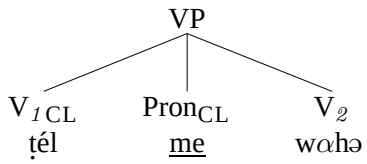


Figure 1: Clitics as postlexical elements

However, there is a group of verbs within class II which do not contain any identifiable derivational prefix.

(18) a. **imperfective**
bɑ̀ylodə́ me
lose I
'I was losing (it).'

b. **perfective**
bɑ́y me lodə
lose$_1$- I -lose$_2$
'I lost (it).' (Tegey 1977, 93)

That is, the element after which the clitic is placed (in the above example *'bɑ̀y'*) does not constitute a morpheme with a separate meaning. It is therefore rather difficult to argue in favor of a clitic status of *'bɑ̀y'* as in Figure 1, if the morpheme is not identifiable as such and furthermore holds a unique position within the language, i.e. it cannot be found in any other word.

3.3 The Special Class of A-initial Verbs

Apart from the three classes introduced above, there is small group of verbs that can have *alternating* stress in the imperfective, but form the perfective with the perfective prefix of class I (*wə-*), thus adopting properties of all three classes. Within this group, there are verbs that begin with consonants, which do not show any special behavior in the imperfective: even if the stress is on the front vowel, the clitic is placed after the verb.

However, there is a small number of verbs in this group with an initial vowel *a-* which show a very distinct behavior with regard to the alternating stress shift in the imperfective. If the stress falls on the second foot, the clitic is placed after the verb ((19a)). If it falls on the initial vowel *a-* however, the clitic is placed directly after the vowel as in (19b), thus acting like the class of bimorphemic verbs.

(19) a. **imperfective — stress on the *second* foot**
ağustə́ me
wear I
'I was wearing it.' (Tegey 1977, 89)

b. **imperfective — stress on the *first* foot**
á me ğustə
wear$_1$- I -wear$_2$
'I was wearing it.' (Tegey 1977, 89)

Apart from the group of verbs discussed in example (18), these *a*-initial verbs are of special interest, because they cannot be clearly identified as bimorphemic verbs and thus display "real" endoclisis. It has been argued that the *a*- was a prefix/clitic (Kaisse 1981, Anderson 2005) from a diachronic perspective, but this cannot be confirmed for all *a*-verbs[4] — furthermore, syncronically, the initial *a*- does not have a recognizable prefix/morpheme-function, as Tegey explicitly states in his thesis (Tegey 1977, 89). The same can be said of the remainder of each form — *ǧust* and all other "remaining" roots are not identifiable as separate morphemes. Hence, additionally to the group of class II verbs where the clitic is inserted after a morphologically unidentifiable item (as in (18)), we have another group of verbs that poses a problem[5] to a postlexical analysis as in Figure 1 and thus seems to violate the Principle of Lexical Integrity.

4 The Postlexical Status of the Clitics

Instead of assuming a postlexical analysis as in Figure 1, another option would be to consider the clitic as being generated in the lexicon, as a part of the morphological word itself, thus preventing the violation. However, there is evidence supporting the fact that the clitic is inserted into the verb *postlexically*. As has been mentioned before, the *a*-initial verbs take the perfective prefix *wə*- like class I verbs. In contrast to the consonant-initial verbs, however, perfective *a*-verbs display *vowel coalescence*, a process that is part of Lexical Phonology (see the overview in Spencer (1996)). In example (20a), the adjacency of the perfective prefix *wə*- and the initial *a*- results in a fusion: *wα*-. In the event of clitic insertion after the perfective prefix, the fused vowel is still present ((20b)), providing evidence that the clitic has been inserted into the word postlexically, that is: after the lexical process of vowel coalescence.

(20) a. **Vowel coalescence — without clitic**
tə ye wαxla (*wə axla)
you it PERF.buy
'You buy it.' (Tegey 1977, 149)

b. **Vowel coalescence — with clitic**
wα ye xla
PERF.buy$_1$ it buy$_2$
'Buy it.' (Tegey 1977, 163)

Given the above examples, Tegey argues that clitic placement takes place after the process of vowel coalescence:

[4]The verbs, where the *a*-vowel cannot be identified as a prefix diachronically are claimed to have been reanalysed (Kaisse 1981).

[5]In that one cannot assign separate syntactic nodes to the two parts of the verb and the clitic.

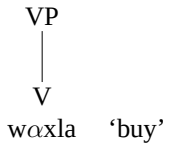
data				**process**
wɔ́	axla		... ye	
wά	xla		... ye	vowel coalescence
wά	ye	xla		clitic placement

Table 2: Tegey's approach: vowel coalescence before clitic placement

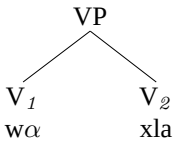
Another argument supporting the analysis of clitics as postlexical elements is that these clitics do not only occur in the context of endoclisis, but act as normal 2P clitics as described in section 2. These clitics exhibit a low degree of selection with respect to their hosts, thus fullfilling a major criterion for a postlexical status (Zwicky and Pullum 1983). The only positional requirement these clitics have to fulfill is the second position. The host of this second position, however, can vary between syntactic constituents and prosodic units as described above. Thus, nobody would question the postlexical status of the 2P *en*clitics — and it seems to be peculiar to describe the same set of clitics as postlexical elements in one context and as lexical in another context, especially since the lexical *endo*clitics can be described as 2P clitics as well: they follow the first accent-bearing foot.

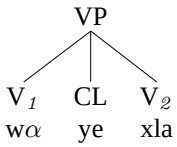
5 Analysis

The lexical process of vowel coalescence described above prohibits the prefix and the verb stem from occupying separate syntactic terminal nodes. That is, they must be viewed as a morphological unit. The monomorphemic verbs thus definitely need to be treated as lexical units and thus can not be represented as in (21b), but are restricted to a representation as in (21a):

(21) a.


'buy'

b.*


c.*


'buy it'

The data on vowel coalescence presented in section 4 leads to the assumption that endoclitics have a postlexical status. Assuming that clitics in general are independent lexical items that should occupy a separate syntactic node leads to a conflict in the case of endoclitics: The integration of the clitic occupying a separate syntactic node would force the verb to split illicitly as demonstrated in (21c). Furthermore, the integration of a postlexical element violates the Principle of Lexical Integrity, which states that a syntactic rule must not interfere with a morphological word.

Numerous approaches to solve this problem have been suggested, but most of them do not provide a satisfactory account of the data. In the following section, the major approaches, their advantages and disadvantages will be discussed to see if some of the findings can be taken as a basis for further development.

5.1 Previous Approaches

5.1.1 Prosodic Inversion

Halpern (1995) proposes Prosodic Inversion (PI) for Pashto (and 2P clitics in general). Halpern assumes a basic underlying (syntactic) structure where the enclitic is swapped with the next available host to its right if no host to the left is provided. This analysis is convenient for a theory like LFG, because it allows functional information to be gathered before the clitic is moved into its prosodically determined position (see e.g. Austin and Bresnan 1996, Nordlinger 1998, Bögel et al. 2010). However, the question remains how the clitic ended up in its syntactic position in the first place. Furthermore, in the specific case of Pashto endoclitics, the violation of the Principle of Lexical Integrity is still given in that the clitic still moves into the word. Even though Halpern assumes PI for Pashto endoclitics, describing them as subcategorizing for a metrical foot, he does not comment upon the problem of Lexical Integrity. Thus, PI might be a possible approach if the above mentioned issues can be resolved, but it does not resolve the architectural issues by itself.

5.1.2 A Different View of Architecture

Kaisse (1981), working within generative grammar, attempts to solve the architectural problem by stating that no phonological rule should precede a syntactic rule, thus assuming that the phonological component is placed after the syntactic component. Kaisse views stress assignment as part of the morphological component, marking a category with phonological information in a first step. In a second step, the clitic moves into its position via syntactic movement rules. It is after the clitic placement that the phonological process of *vowel coalescence* takes place. Kaisse thus assumes an architecture like the following:

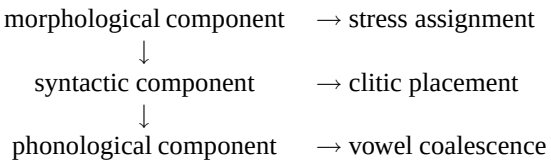


Figure 2: Kaisse's architectural assumption

This architectural view stands in contrast to the assumption that syntax and prosody form two parallel and interacting, but independent modules (e.g. Inkelas and Zec 1990). Furthermore, it does not provide a satisfactory account of the endoclitic phenomenon. In order to avoid the problem of endoclitics, Kaisse claims that Pashto displays no real endoclisis by arguing that all Pashto verbs allowing endoclisis can be described as bimorphemic i.e. as containing a prefix after which the clitic attaches. However, even if this could be verified (but see the argumentation in section 3.2. and 3.3.), the prefixes would still be part of the morphological word — thus there would still be a violation in this approach.

5.1.3 Optimality Theory

van der Leeuw (1997), Roberts (1997) and Anderson (2005) analyse Pashto endoclitics within Optimality Theory (OT) (see e.g. McCarthy 2001). Roberts states that Pashto endoclitics should be viewed in phonological terms only, although he also assumes that phonological phrases are derived from maximal syntactic projections (along the lines of Selkirk (1986)). In his approach principles of OT select the output form. Anderson also assumes the OT constraint ranking ((22)):

(22) Integrity(DP), Integrity(PP), Integrity(PPhrase), NonInitial(cl_i, IP) $>>$ LeftMost(cl_i, IP)

(Anderson 2005, 154)

which reads as: "The clitic is oriented towards the left edge of the IP; however, it must not appear in the initial position and the integrity of the DP, the PP and the Phonological Phrase[6] appearing in this initial position must be preserved". Thus, Anderson assumes the phonological phrase to be the prosodic host for the clitics. He follows Kaisse in that he dismisses the existence of endoclitics, but views all *a*-initial verbs as complex verbs as well.[7] In his approach the status of a phonological word is assigned to any lexical element that bears stress. Phonological phrases are then constructed on the basis of phonological words — allowing phonological phrases to consist of only one phonological word as well. The question remains open as to how this initial verb sequence is analysed in prosodic terms — to consider this first element a "phonological word" or even a "phonological phrase" (Roberts 1997) seems odd, particularly if not even the prefixal status is confirmed (*a*- initial verbs and subgroup of class II verbs). However, even assigning an independent phonological word status to prefixes seems disproportionate, especially since it looses this status so easily if the stress is on the second part of the verb. As an alternative, I suggest that this initial element should be described as "the foot bearing main stress" (along the lines with Kopris and Davis (2005) and Halpern (1995)).

	Roberts (1997)	**Anderson (2005)**	**New suggestion**
prosodic unit	Phonological Phrase	Phonological Word/Phrase	Foot
example	(P-Phr)-cl (P-phr) á me xistələ	(P-Wrd)-cl (P-Wrd) á me xistələ	(x)-cl (x x) á me xistəl(ə)

Table 3: Determining the prosodic host

However, not all approaches to clitics share the belief in prosody as the driving factor. In a later account of Pashto clitics, Roberts (2000) discards his former prosodic approach and claims that the clitics are actually agreement morphemes

[6] Thus accounting for examples like (9), where the unstressed material is ignored by the clitic.

[7] Note that Anderson does not comment on the group of class II verbs that cannot be analysed as complex verbs, but allow endoclitics ((18)). Kaisse mentions them briefly, but in a separate context.

merged into a high position in the clause, reducing the phonological operations to a minimum. The only operation Roberts considers as a "last resort" is Prosodic Inversion to explain the endoclitics.

5.1.4 Word Order Domains/ Topological Fields

Dost (2005) argues strongly against Roberts and proposes an interaction of syntax and prosody to be involved in clitic placement. His approach is based on word order domains/topological fields in combination with a Head-driven Phrase Structure Grammar (HPSG) architecture (Pollard and Sag 1994). His distinction between vertical hierarchy and linear precedence is along the lines of Bögel et al. (2010), but he does not support the idea of Prosodic Inversion and suggests instead that even though the verbs consist of *one* syntactic atom, they remain separable in terms of the word order domain; that is, these verbs contain more than one domain and are thus internally complex. However, the architecture neither makes reference to word-internal feet or stress nor (as Dost himself points out) does it give a full account of what the interaction between the prosodic and the syntactic component should look like. Further research needs to be done in that direction. Dost also emphasizes the fact that the Principle of Lexical Integrity is not violated in a strict sense in that he does not apply *syntactic* processes to interfere with the internal structure of the word, but views clitic placement as resulting from the interaction of syntax, prosody and word order domains — an idea fully supported by this paper.

5.1.5 Lexical Sharing

Another approach that has been lately debated within the LFG community is Lexical Sharing (Wescoat 2002). Lexical Sharing is an application allowing two terminal nodes in the c-structure to share one lexical item. To achieve this, Wescoat assumes that each word is separated from its terminal node and put into a linearly ordered set, the *l-structure*. Wescoat then introduces a structural correspondence between c- and l-structure in the form of a *lexical exponent mapping* λ that generally is a one-to-one mapping between the terminal node and the word, but also allows for two or more terminal nodes to refer to one word. The following formula refers to the f-structure of the lexical component of the current node:

(23) $\varphi(\lambda(*))$, in short: $\Downarrow$

where λ represents the mapping from *l-* to *c-structure* and φ the mapping from *c-* to *f-structure*. A short description of the basic components of the theory is given in Table 4:

l-structure	→ lexical-exponence rules, contribute an independent set of functional descriptions
λ	→ maps the words to the terminal nodes of c-strucure
c-structure	→ two syntactically aligned terminal nodes may share one lexical item
φ	→ transports functional information to f-structure

Table 4: The basic features of Lexical Sharing

Udi Person Markers

Wescoat (2009) applies this theory to Udi person markers (Harris 2002), another form of endoclitics. He views the Udi person markers as "instantiation-altering morphemes" — a word containing a person marker instantiates two terminal nodes. Such an application would, of course, solve the problem posed by the Pashto endoclitics discussed here. Thus, this section takes a close look at Wescoat's application of Lexical Sharing to Udi person markers.

According to his theory, a word can be the lexical exponent of two terminal nodes. Wescoat gives several examples of what these lexical exponence rules should look like; in principle they are ordered as in example (24), where the Udi verb *beɣ-al* 'watch' is combined with an enclitic *le*:

(24) beɣ-al-**<u>le</u>** ←

V	PM
(↓ PRED) = 'WATCH<(↓ SUBJ),(↓ OBJ)>'	(↑ SUBJ) = ↓
(↓ TNS) = FUT	(↓ PERS) = 3
⇓ = ↓	(↓ NMB) = SG

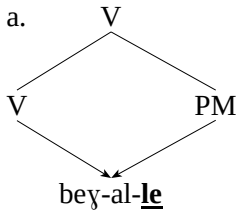
(Wescoat 2009)

Generally clitics are viewed as independent syntactic items that are phonologically dependent on a host; however, Wescoat uses the term "morpheme" to describe the person markers and refers to Harris' alignment constraints based on Optimality Theory to explain the exact position of these clitics. Even though his argumentation seems to be quite straightforward, the question remains whether he views these person markers as being generated in the lexicon as part of the word or as being attached to the word in a later process. If the former is true, then this process would be highly inefficient, since these "morphemes" must be allowed to attach to a large variety of hosts. On the other hand, following the latter assumption, if the clitic is attached to the word later on, it is unclear how it acquired this position, let alone the position *within* the word, without violating the Principle of Lexical Integrity. Still, as mentioned before, Wescoat's analysis is interesting in that it allows two nodes to correspond to one lexical item. Thus I will pursue his appoach a bit further and take a close look on what happens within the syntax.

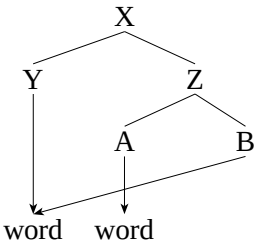
The terminal node PM corresponding to the person marker, which is aligned with the verb by morphological alignment constraints and thus associated with the terminal node V in the lexical exponence rule ((24)), is positioned by syntactic

constraints. It shares a mother node with the terminal node of the verb, leading to a representation as in (25a). The relative order of the terminal nodes must be preserved, avoiding representations as in (25b), where two nodes correspond to one lexical item crossing another lexical item:

(25) a.



*b.



According to his analysis, even though Udi person markers can be found in all kinds of positions in the sentence, their constituent can be adjoined to either S, VP or V, thus allowing for the functional information of the marker to be picked up by the f-structure. Even though it is not deeply relevant to the analysis, it remains unclear how this syntactic placement of the PM terminal node would be treated within c-structure in the case of endoclisis: is it the position preceding or following the verb that is reserved for the terminal node of the endoclitic? And on what grounds is this decision made?

Pashto clitics and Lexical Sharing

After this brief survey of the implementation of Udi person markers, I now turn to Pashto to see if Lexical Sharing can be applied to this phenomenon as well, solving the main problem of allowing two syntactic nodes corresponding to one lexical item.

A lexical exponence rule as used for Udi person markers in (24) could be postulated for Pashto endoclitics as in (26).

(26) verb$_1$ - clitic -verb$_2$ ← V CL
... ...

Assuming that the words including the clitics have to be generated in the lexicon, this approach will lead to an immense number of forms within the lexicon:

1. *Each verb will have to be listed with each clitic as* endo- *and* en*clitic*: As we can see in section 2, the group of clitics discussed in this paper involves 10 different clitics, which would all have to be listed.

2. *Each non-verbal host in the language has to be listed with these clitics as well*: Pashto clitics and second position clitics in general show a high promiscuity with respect to their host; thus, approximately half of the words in the language can act as potential hosts and must be encoded as such within the lexicon.

3. *Each of the word-clitic combinations under 1. and 2. can be combined with any other clitic*: All of the clitics in question can cooccur, thus all possible cooccurences have to be listed in the lexicon as well

The required listing of clitics and their hosts in the lexicon would be a very unsatisfying and inelegant solution. And even if we ignore this and continue with a syntactic representation, we encounter further problems. Since these clitics can possibly cooccur with each other, each one of them can instantiate an independent syntactic node. Given this, one has to consider the template in (3), which forces a certain order on the clitics in question. Using the corresponding terminal nodes, this template would look like (27):

(27)

ADV	MOD	PRON	PRON	PRON	PRON/MOD	PRON	ADV
1	2	3	4	5	6	7	8
xo	ba	am	am/mo	me	de	ye	no

If several clitics are involved in an expression, the order in (27) would have to be maintained in the instantiated terminal nodes. However, it is hard to imagine how the order of terminal nodes in the first line of (27) could be justified within syntax. If syntactic constraints were involved we would not encounter an adverb at the beginning and at the end interrupted by numerous pronouns and modals as sketched in Figure 3:[8]

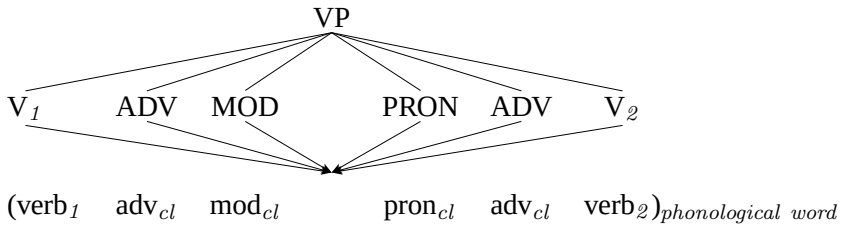


Figure 3: Multiple endoclitics in a verb represented by Lexical Sharing (sketch)

I conclude that the terminal nodes "sharing" one lexical item are obviously not internally ordered by syntactic constraints[9] — their position in the syntax is difficult to justify. This goes along with the fact that the positioning of Pashto (endo)clitics is regulated by prosody, a component that has not been dealt with (yet) within Lexical Sharing.

To sum up, although Lexical Sharing is one possible way of representing one lexical item with two terminal nodes, it cannot be applied to Pashto (endo)clitics for the reasons listed above.

[8] A reordering of the nodes is excluded as well, since this might lead towards the direction of representations as in (25b).

[9] However, if the order of the clitics within the template is on the basis of prosodic/phonological constraints or if the order is 'accidental' is difficult to say. One also could assume that the syntax has a special category for these clitics, e.g. CL_1, CL_2... In this case, the ordering would not be an issue for the c-structure placement.

5.2 Proposed Approach

The evaluation presented above thus excludes certain possibilities of dealing with endoclisis:

1. The clitics and all possible hosts have to be listed in the lexicon — an unsatisfying and inelegant solution.

2. The combination of the word and the clitic cannot instantiate two or more terminal nodes, because the internal ordering of the clitics' terminal nodes is not syntactically justified.

Thus, other approaches have to be taken into consideration. One possible solution is the separation of the linear order of prosodic and syntactic elements. This has been suggested before by proposals like Prosodic Inversion (see section 5.1.1.) and specifically within LFG recently by Bögel et al. (2010). In this approach, the syntactic representation includes the clitics in the first position of the sentence, thus gathering the information for the f-structure from this location. The prosodic representation then determines the position of the clitics as pronounced in an utterance via the application of Prosodic Inversion, placing the clitic in the correct second position as determined by syntactic or prosodic constraints.

However, this approach faces the same problem concerning lexical integrity as this paper in that the movement of a clitic *into* a word causes a violation of the Principle of Lexical Integrity. Thus, a solution to this problem could also be seen as an extension to this approach. In general, it would be desirable to represent the actual prosodic succession of elements in the syntax as well as securing functional information. However, such an approach does not (yet) exist — further research needs to be done. Generally though, the basic concept of viewing prosody and syntax as two independent but interacting components would seem to be just right for Pashto endoclitics.

Hock (1996), also viewing prosody as decoupled from syntax, states that second position clitics should be placed after the first "accented element" of an inital verbal clitic group and after the first accent-bearing constituent elsewhere. With these constraints, one can account for all of the above Pashto examples. Thus, the requirement of a clitic to attach to a host is a strong prosodic requirement. In an architecture that assumes syntax and prosody as interacting, but decoupled dimensions, the placement of Pashto endoclitics can be explained due to prosodic (and not syntactic) constraints. This assumption is represented in the following (rough) architecture, showing the prosodic effects in a parallel architecture with a class II verb ((28)):

(28) ṭél me wɑhə
PREF I push
'I pushed (it).' (Tegey 1977, 92)

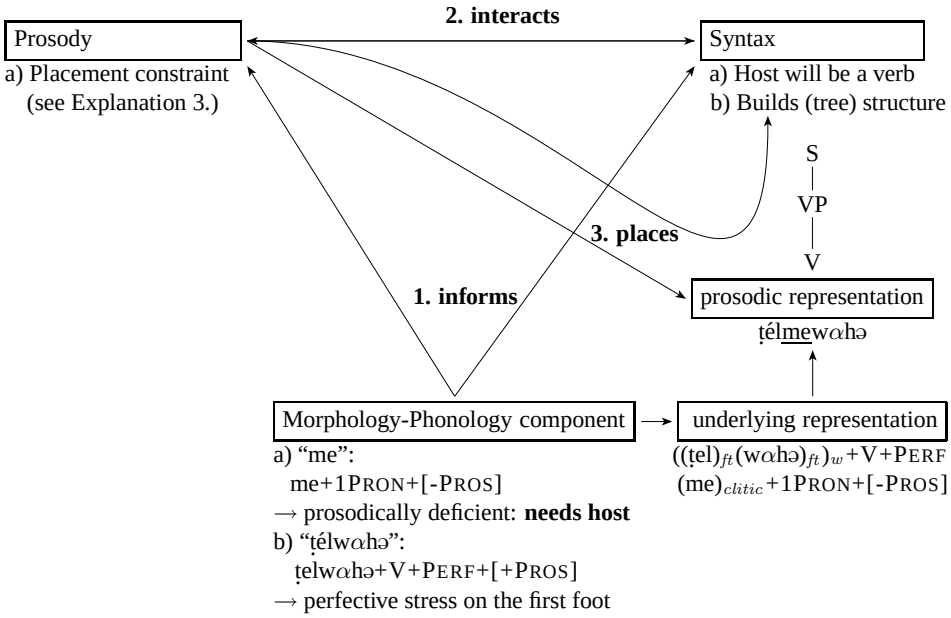


Figure 4: A representation of the parallel architecture

Both lexical items, the verb *ṭelwαhə* 'push' and the clitic *me* 'I' are analysed within the MORPHOLOGY-PHONOLOGY component. Here, the verb receives perfective stress on the first foot because of the perfective aspect ([+PERF]) and is indicated as forming a prosodic word ([+PROS]); the clitic on the other hand is marked as prosodically deficient ([−PROS]). The component then (1.) **informs** syntax and prosody of the properties of the lexical items, i.e. the analysis of the words. The clitic is recognized as a prosodically deficient item ([−PROS]). It needs to be attached to a host. The functional and phonological information of each lexical item is stored in the *underlying representation*. PROSODY and SYNTAX (2.) **interact** in that they share information on structure and intonation. Syntax provides the information that it will be the verb that has to be the host of the clitic. PROSODY is responsible for the placement of the prosodically deficient item. Depending on the host, the clitic is palced in a certain position. Since the host is a verb, the placement is subject to a prosodic constraint: the clitic has to be placed after the first stressed foot of the verb. Prosody thus (3.) **places** the clitic after the first stressed foot of the host, which leads to the *prosodic surface representation*: *ṭélmewαhə*.

Thus, the above architecture tries to represent the "parallel" approach in which a morphological component interfaces with the LFG syntax and an independent prosodic representation. These three components are each independent and governed by independent rules and principles. However, they must also interact and the complex nature of their interaction is brought out nicely by phenomenon such as the Pashto endoclitics, allowing prosody to interact with the lexical word and thus not violate the Principle of Lexical Integrity.

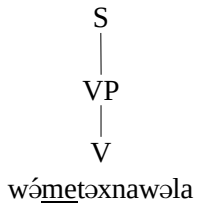
(29) PROSODIC INTERACTION PRINCIPLE:

While syntax may not intervene in the word-internal structure after the morphological word is formed, prosody still has access to the internal structure of the prosodic word (e.g. the footing).

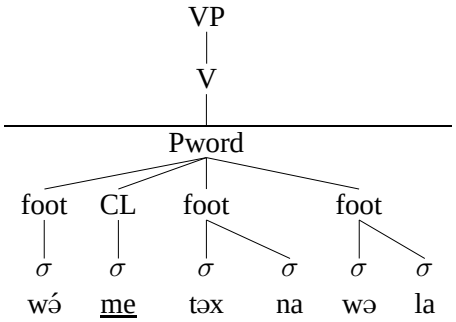
If it is the case that there is no host where a prosodically deficient item can attach to at the right edge, the prosody has the power to overrule the Principle of Lexical Integrity and place the clitic appropriately according to the prosodic structure. In Pashto, this would be after the first accent-bearing foot in a verb-initial clause, leading to a syntactic representation as in (30b) and a prosodic representation as in (30c), where the clitic is inserted after the first foot bearing main accent.

(30) a) wə́ me təxnawəla
PREF I tickle
'I tickled (her).'

b)



c) Proposed tree with syntactic and prosodic structure of (30a).



6 Conclusion

This paper presents the problem of Pashto endoclitics, which challenge established views of the prosody-syntax interface and notions of lexical integrity. Different approaches are evaluated and conclusions are drawn from these evaluations. The result is the proposal of a solution involving an architecture of grammar in which morphology, syntax and prosody are taken to be independent, but interacting modules of grammar. As much as possible, the three components align with one another; however misalignments are also allowed. In particular, prosody is allowed to misalign with syntax when a prosodically deficient item like a clitic needs to be placed in a prosodically appropriate position. In the case of Pashto endoclitics, this is after the first accent-bearing foot. Thus, postlexical prosodic requirements are taken to allow for the placement of material into a morphologically well-formed

and complete word, thus evading a violation of the Principle of Lexical Integrity. However, this paper is only a first proposal in this direction — further research is necessary, especially on the exact nature of the prosody-syntax interface.

References

Anderson, Stephen. 2005. *Aspects of the Theory of Clitics*. Oxford University Press.

Austin, Peter and Bresnan, Joan. 1996. Non-Configurationality in Australian Aboriginal Languages. *Natural Language and Linguistic Theory* 14(2).

Bögel, Tina, Butt, Miriam, Kaplan, Ronald M., King, Tracy Holloway and Maxwell III, John T. 2009. Prosodic Phonology in LFG: a New Proposal. In *Proceedings of LFG09*, 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 *Proceedings of LFG10*, CSLI Publications.

Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Blackwell.

Bresnan, Joan and Mchombo, Sam. 1995. The Lexical Integrity Principle: Evidence from Bantu. *Natural language and Linguistic Theory* 13(2), 181–254.

Butt, Miriam. 2007. The Role of Pronominal Suffixes in Punjabi. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling and Chris Manning (eds.), *Architecture, Rules, and Preferences*, pages 341–368, CSLI Publications.

Dost, Ascander. 2005. A Domain-Based Approach to 2P Clitics in Pashto. In *Proceedings of the Texas Linguistics Society IX Conference*, CSLI Publications.

Franks, Steven and King, Tracy Holloway. 2000. *A Handbook of Slavic Clitics*. Oxford University Press.

Halpern, Aaron. 1995. *On the Placement and Morphology of Clitics*. CSLI Publication.

Halpern, Aaron and Zwicky, Arnold (eds.). 1996. *Approaching Second: Second Position Clitics and Related Phenomena*. CSLI Publications.

Harris, Alice C. 2002. *Endoclitics and the Origins of Udi Morphosyntax*. Oxford University Press.

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

Inkelas, Sharon and Zec, Draga (eds.). 1990. *The Phonology-Syntax Connection*. CSLI Publications.

Jackendoff, Ray. 2010. *Meaning and the Lexicon*. Oxford University Press.

Kaisse, Ellen M. 1981. Separating Phonology from Syntax: A Reanalysis of Pashto Cliticization. *Journal of Linguistics* 17, 197–208.

Kopris, Craig A. and Davis, Anthony R. 2005. *Endoclitics in Pashto: Implications for Lexical Integrity*. Presented at the Fifth Mediterranean Morphology Meeting, Sept 15-18, Fréjus, France.

Lahiri, Aditi and Plank, Frans. 2009. What Linguistics Universals can be true of. In Sergio Scalise, Elisabetta Magni and Antonietta Bisetto (eds.), *Universals of Language Today*, pages 31–58, Springer.

McCarthy, John. 2001. *A Thematic Guide To Optimality Theory*. Cambridge University Press.

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

Nordlinger, Rachel. 1998. *Constructive Case: Evidence from Australian Languages*. CSLI Publications.

Pollard, Carl and Sag, Ivan. 1994. *Head-driven Phrase Structure Grammar*. University of Chicago.

Roberts, Taylor. 1997. The Optimal Second Position in Pashto. In Geert Booij and Jeroen van de Weijer (eds.), *Phonology in Progress - Progress in Phonology: HIL Phonology Papers III*, pages 367–401, Holland Academic Press.

Roberts, Taylor. 2000. *Clitics and Agreement*. Ph. D. thesis, Massachusetts Institute of Technology.

Selkirk, Elisabeth. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. MIT Press.

Selkirk, Elisabeth. 1986. On Derived Domains in Sentence Phonology. *Phonology Yearbook* 3, 371–405.

Spencer, Andrew. 1996. *Phonology*. Blackwell.

Tegey, Habibullah. 1977. *The Grammar of Clitics: Evidence from Pashto and Other Languages*. Ph. D. thesis, University of Illinois.

van der Leeuw, Frank. 1997. *Clitics: Prosodic Studies*. Holland Academic Graphics.

Wescoat, Michael T. 2002. *On Lexical Sharing*. Ph. D. thesis, Stanford University.

Wescoat, Michael T. 2009. Udi Person Markers and Lexical Integrity. In *Proceedings of LFG09*, CSLI Publications.

Zec, Draga and Inkelas, Sharon. 1990. Prosodically Constrained Syntax. In Sharon Inkelas and Draga Zec (eds.), *The Phonology-Syntax Connection*, pages 365–378, CSLI Publications.

Zwicky, Arnold and Pullum, Geoffrey K. 1983. Clitization vs. Inflection: English *n't*. *Language* 59, 502–513.