

In-cliticization? Apparent word-internal clitic placement in Pashto and Old Irish

1 Introduction

The placement of second-position clitics in Pashto has received some attention in the literature because of the complications it poses for what it means to be in “first” or “second” position. Different languages define these requirements in different ways: in Ancient Greek for example, as noted by Wackernagel (1892), second-position clitics occur regularly after the first word in a clause, as in (1). Clitics are marked in boldface type and underlined in examples throughout this paper.

- (1) [polees]_ω **te min** ērēsanto hippēes phoreein (Iliad 4.143)
many and it prayed riders carry
'And many riders prayed to carry it.'

In languages of this sort, second position can be defined, without reference to syntactic constituents or even to specific prosodic demands, as the position following the first word. Another possible definition for second position is available as an option in Serbo-Croatian, where a second-position clitic may follow either the first word as in Ancient Greek (2b), or following the first syntactic constituent (2a):

- (2)a. [_{DP} Lav Tolstoi] **je** veliki ruski pisac (Halpern, 1995, via Roberts 1997:380)
Leo Tolstoy is great Russian writer
- b. [Lav]_ω **je** Tolstoi veliki ruski pisac
Leo is Tolstoy great Russian writer
'Leo Tolstoy is a great Russian writer.'

In (2), it appears that the clitic is able to choose between looking at the syntax and looking at the prosody in finding second position in its clause, which provides some of the impetus for syntax-based analyses of clitic placement. However, a purely prosodic approach would work here as well. Assuming the DP is mapped onto a phonological phrase, the optionality available to the clitic becomes not a matter of syntax versus phonology, but whether to attach following the first phonological word or first phonological phrase.

Pashto shows what appears to be a similar pattern of variation between having first position filled by either a phonological word or a phonological phrase, but without the freedom to choose between the two in any given sentence. The sentences in (3) exemplify some of the range of this pattern. (3a) shows a clitic attaching to a phonological phrase, in (3b) the clitic attaches to a phonological word, and in (3c) it seems to attach to a morpheme within a word.

- (3)a. [_{NP} aʎa ʃel kaləna danga aw khaysta peyla] me nən bya wəlida (Tegey 1977:83-4)
that 20 year tall and pretty girl I today again saw
'I saw that twenty-year-old tall and pretty girl again today.'
- b. [_{VP} axistələ] me (Tegey 1977:89)
buy I
'I was buying them.'
- c. [ʔ á] me xistələ
PREFIX I buy
'I was buying them.'

Purely or mostly syntactic approaches to the question of clitic placement have been attempted (see Roberts (2000) and Kaisse (1981) for Pashto), as have mixed syntax-prosody accounts using inversion (see Halpern (1995) for Serbo-Croatian, and a brief look at this in Pashto in Roberts (1997)), though neither seems to explain all of the data. Purely phonological analyses include Roberts (1997) and Dost (2007), who has an HPSG account. These have had more success, but to carry the analysis as far as is necessary to explain data such as in (3c), they rely on a typologically unusual prosody involving a prosodic phrase consisting of a single vowel, and a clitic appearing in the middle of what is apparently a single phonological word. In this, comparison with Old Irish pronominal clitics can provide some support: in Old Irish, second-position clitics (often referred to as infixes in the grammars (Thurneysen 1961:255)) can appear between a monosyllabic preverbal particle and the verb itself.

- (4)a. Fo -gaib in duine eich ón chennaigiu¹ (Quin 1975: 40-45)
PREVERB-get.3SG the man horses from.the merchants.DAT
'The man gets horses from the merchants.'
- b. Fo -s -gaib in duine ón chennaigiu
PREVERB-3PL-get.3SG the man from.the merchants.DAT
'The man gets them from the merchants.'

Although Old Irish verbs which contain a preverb, like *fo-gaib*, are written with a hyphen in grammars, this was not the case in manuscripts, and there is no reason to think that the monks who wrote them thought of these verbs as anything other than one word.

Another parallel to Pashto is that the placement of these clitics in Old Irish also shows a stress alternation. When the Old Irish verb is initial in its clause, as in (4a-b), the main stress falls on the first syllable following the preverb. However, when the verb is preceded by a negation or question particle, the main stress of the verb falls on the preverb, and the enclitic appears directly following the particle. The shift in stress also causes phonological/morphological changes to occur in the verb.

1 In Old Irish, an acute accent marks a long vowel. When necessary, stress will be marked with the IPA symbols for it.

- (5)a. ˌFo -'gaib eich
PREVERB-get.3SG horses
'He gets horses.'
- b. ˌFo -s -'gaib
PREVERB-3PL-get.3SG
'He gets them.'
- c. ˌNí -s 'fagaib
NEG-3PL get.3SG
'He does not get them.'
- d. *ˌNí 'fa -s -gaib
NEG PREVERB-3PL-get-3sg
- e. *ˌNí 'fagaib -s
NEG get.3SG-3PL

In (5a), the preverb and verb appear together, with stress on the second syllable. When the overt object is replaced by a clitic pronoun in (5b), it appears between the preverb and the verb. However, when negation appears before the verb, the preverb becomes stressed, and the enclitic appears after the negation, rather than after the stressed preverb (5d) or whole verb (5e).

In this paper, I attempt to show that clitic placement in both Pashto and Old Irish can be explained with a phonological account, and that evidence from these two languages can give us a picture of what the syntax which feeds the phonology looks like. In section 2, I present the language data in more detail, highlighting the problems which need to be accounted for. In section 3, I propose syntactic structures for these sentences. In section 4, I look at how the syntax-prosody mapping works. Section 5 is an Optimality Theoretic account of clitic placement in not only the Pashto/Old Irish-type languages, but also for Ancient Greek and Serbo-Croatian. Section 6 is a conclusion.

2 Clitics in Pashto and Old Irish

2.1 Pashto

Pashto is an Indo-Iranian language spoken in Afghanistan. It has SOV word order and is quite rigidly verb-final, compared to Persian, which allows postposed post-verbal PP complements. This paper focuses on second position clitics, but general information on Pashto grammar can be found in Tegey & Robson (1996) and Babrakzai (1999) along with additional data relevant to this topic.

2.1.1 Pashto Clitics

Pashto has two kinds of clitics: second position clitics, and verbal clitics. This paper only deals

with the former, which include pronominals, modals, and adverbials. These are listed below (Tegey, 1977:81):

(6) Second-position clitics:

Pronominal (ergative, accusative, genitive)

me 1SG
de 2SG
ye 3SG, 3PL
am 1PL, 2PL
mo 1PL, 2PL

Modal

ba 'will', 'might', 'must', 'should', 'may'
de 'should', 'had better', 'let'

Adverbial

xo 'indeed', 'really', 'of course'
no 'then'

When these co-occur, they appear in the following order (Tegey 1977):

(7)	1	2	3	4	5	6	7	8
	xo	ba	am	am,mo	me	de	ye	no
	'indeed'	'will'	1PL,2PL	1PL,2PL	1SG	2SG,'should'	3SG,3PL	'then'

The following group of examples shows for the clitic *de* 'should' that it always appears in second position in the sentence as each initial item is removed:

- (8)a. tor **de** nən xar nə rawali
Tor should today donkey not bring
'Tor should not bring the donkey today.'
- b. nən **de** xar nə rawali
today should donkey not bring
'He should not bring the donkey today.'
- c. xar **de** nə rawali
donkey should not bring
'He should not bring the donkey.'
- d. nə **de** rawali
not should bring
'He should not bring it.'

- e. rawali **de**
bring should
'He should bring it.'

Roberts (1997:370) points out that (8e) is of particular interest because Pashto is so rigidly verb-final and that "it illustrates that the clitic's need to have a host to its left is strong enough that it compels the verb to appear non-finally in a sentence containing only one word (the verb) other than the clitic." I would argue that rather than evidence for a constraint-based approach that favors enclitics leaning left over verb-finality, that this example shows that these clitics are completely post-syntactic. By the time they are placed, the syntax has already been fed into the prosody, and any preference for a verb-final structure is no longer relevant.

However, as we saw an extreme example of in (3a) (repeated below as (9), these clitics do not look strictly for the first word in a clause to attach to:

- (9) [_{NP} aʎa ʃel kaləna danga aw khaysta peyla] **me** nən bya wəlida (Tegey 1977:83-4)
that 20 year tall and pretty girl I today again saw
'I saw that twenty-year-old tall and pretty girl again today.'

Instead, they attach following the first constituent, exemplified again in (10):

- (10) [_{DP} xuʃal aw patang] **be ye** dər ta rawɽi (Tegey 1977:83-4)
Khosal and Patang will it you to bring
'Khosal and Patang will bring it to you.'

2.1.2 Pashto Verbs

Pashto verbs are divided into three classes and one sub-class, which differ in how they form their imperfect and perfect. In every case, the difference between perfect and imperfect is marked by a shift in stress. In class I are monomorphemic verbs and form their perfect with a particle wá, which takes the main stress off the verb. (Though these verbs can appear optionally with initial or non-initial stress.) The sub-class is a set of nine apparently monomorphemic verbs which begin with a-. These also take wá to form the imperfect, but it merges phonologically with the a-, and appears as wá, separate from the rest of the verb. (These also show optional initial stress, which causes an enclitic to follow the initial a-, as in (3c).) The second class contains verbs which consist of a prefix+root. To form the perfect for these, the stress shifts from the root onto the prefix. In the third class are verbs formed with a compound lexical item+auxiliary verb. To form the perfect in these, the stress shifts from the auxiliary onto the lexical item and the form of the verb changes. (Tegey 1977, via Kaisse 1981.)

(11)		Imperfect		Perfect
Class I				
Monomorphemic				
skundəla 'pinch' (past)	skundála	'was pinching'	wá skundəla	'pinched'
gadədə 'dance' (past)	gadədə́	'was dancing'	wá gadədə	'danced'
a- initial				
axistəl 'to buy'	axistələ́	'was buying'	wá xistələ	'bought'
aɣustəl 'to wear'	aɣustə́	'was wearing'	wá ɣustə	'wore'
Class II				
Compound prefix+root				
tel-wahə 'push'	telwahə́	'was pushing'	télwahə	'pushed'
pore-westə 'carry across'	porewestə́	'was carrying across'	pórewestə	'carried across'
Class III				
Compound lexical item+auxiliary verb				
wobə kawi 'water do'	wobə káwəl	'is watering'	wobə́ krə	'watered'
porta kigi 'up become'	porta kígi	'is getting up'	pórta swa	'got up'

In Class I verbs, the location of stress can be initial or non-initial in the imperfect. In the a-initial sub-class, initial stress causes a clitic to appear following the first syllable, apparently dividing the word:

- (12)a. a-xistələ́ me (Tegey 1977:89)
 buy I
 'I was buying them.'
- b. á me xistələ́
 PREFIX I buy
 'I was buying them.'

This a- has no synchronic meaning, which is the reason Tegey and others have grouped it with the monomorphemic verbs, despite sharing some behaviors with the compound ones. Historically, this comes from an Old Persian augment that was an imperfect marker.

The consonant-initial Class I verbs also show optional stress alternations in the imperfect, but the position of clitics does not change:

- (13) a. *satám ye*
keep it
'I keep it.' (Tegey 1977:88)
- b. *sátəm ye*
keep it
'I keep it.'

Verbs in classes II and III follow the same pattern as the a-initial verbs in that clitics will follow the first element (in a sentence consisting of only a verb) if that element is stressed, but will follow the entire compound if the first element is unstressed. Verbs in class I in the perfect will have clitics following the stressed perfect morpheme, *wá*.

2.1.3 Pashto prepositional phrases

The other major hurdle to a definition of second-position for Pashto clitics is the possibility of unstressed prepositional phrases occurring at the beginning of a sentence. If the prepositional phrase contains a noun which bears stress (as in (14a)), the clitics will appear after the noun, but if it contains only a preposition (as in (14b)), the clitics will appear on the next stressed element. In (14c) we see an extreme example of a second position clitic that has to move quite far to the right to find a suitable host.

- (14)a. [_{PP} pə rasáy] *ba ye* wá-taʔi (Tegey 1977:114)
with rope will it PERF-tie
'He will tie it with the rope.'
- b. [_{PP} pe] wú *ba ye* taʔi
with PERF will it tie
'He will tie it with it.'
- c. [_{PP} ra ta] [_{PP} te] [_{AdvP} ra] ʃkawá *de* (Tegey 1977:119)
me for from-it here pick you
'You were picking it for me from it (and bringing it) here.'

2.1.4 Summary

Pashto second position clitics follow a number of criteria in their placement. They occur following the first prominent stress in a sentence, which looks like an Ancient Greek-style placement when this first stress is a simple noun. However, if it is a complex DP or conjoined DP, the clitics will not break up constituent. If the first element is part of a compound verb, the clitics will appear after the first element of the verb, if it is stressed. Unstressed prepositional phrases can also appear sentence-initially, and seem to be invisible to the clitics in choosing their host.

2.2 Old Irish

Old Irish is the Celtic language spoken in Ireland in (about) the 7th - 9th centuries and is the ancestor of Modern Irish and Scottish Gaelic. It is VSO, but there is a small class of particles which can appear preverbally. These include negation and question markers, some complementizers such as “if”, and *wh*-words, though these take a relative clause following them and do not affect the syntactic position of the verb. The most notable feature that sets Old Irish apart from its descendants is its highly complex verbal system involving two sets of conjugations for each verb: one for when it appears in absolute initial position, and another when it follows a preverbal particle.

2.2.1 Old Irish Verbs

As mentioned above, simple verbs in Old Irish alternate between their clause-initial form (absolute form), and a different form when following a particle. In their 2000 paper, Carnie, Harley & Pyatt (henceforth CHP) analyze this as $V \rightarrow C^0$ movement in the former case and $V \rightarrow \text{Infl}$ movement in the latter. Their basic structure is as follows:

- (15)a. Beirid /Ní beir in fer in claideb. (CHP 2000:46)
 Carry.3SG/NEG carry.3SG the man the sword
 'The man carries/does not carry the sword.'
- b. $[_{CP} \text{Beirid}_i + C^0 [_{IP} t_i [_{VP} \text{in fer} [_{V'} t_i \text{in claideb}]]]]$
- c. $[_{CP} \text{Ní} [_{IP} \text{beir}_i + \text{Infl} [_{VP} \text{in fer} [_{V'} t_i \text{in claideb}]]]]$

CHP suggest that this dual set of conjugations is the result of the verb "incorporating into a null C^0 " (2000:46). When the verb combines with Infl, as in (15c), it appears in what is called its conjunct form, but when the verb raises to C^0 , as in (15b), it takes a different form as a result.

This system of alternating forms for simple verbs also comes into play when forming morphologically complex verbs. These are combined by prefixing a preverbal particle to the conjunct form of a verbal stem. These particles are homophonous with, and presumably derived from, prepositions. This method of forming verbs should be familiar to speakers of most Indo-European languages:

- | | | | |
|--------|---------------------------------|---|----------|
| (16)a. | Scribere
write
'To write' | Con-scribere
with-write
'To enlist' | (Latin) |
| b. | Sehen
see
'To see' | Aus-sehen
out-see
'To appear' | (German) |

Enclitics attach to the first stressed item in its clause in Old Irish. This can be a verb in absolute form as in (19a), a negation particle before a simple verb as in (19b), a preverbal particle as in (19c) or negation before a complex verb as in (19d):

- (19)a. 'Beirth -i
carry.3SG-3SG
'He carries it.' (Strachan 1949:25)
- b. ,Ní -m 'beir
NEG-1SG carry.3SG
'He goes not carry me.'
- c. ,Do -m 'beir
PREVERB-1SG bring.3SG
'He brings me.'
- d. ,Ní -m 'tabair
NEG-1SG bring.3SG
'He does not bring me.'

Although these examples show the enclitics appearing straightforwardly in second position, there are also a few function words which could show up before a preverb, and complicate the pattern:

- (20) ma ,immi -m 'thabarthar, (Milan Glosses, 41^c2)
if PREVERB -1SG surround.3SG.present.deponent
'If I am surrounded,'

2.2.3 Summary of Old Irish

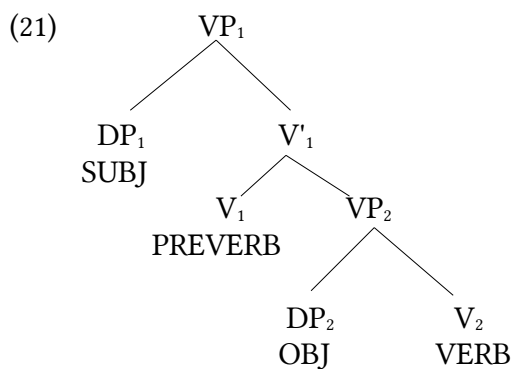
As was mentioned in the introduction, although preverbs are separated from verbal stems by a hyphen in the grammars, they were written without punctuation or spaces between them in the manuscripts, and there is no reason to think that the monks or speakers of Old Irish thought of *dobeir* as any less one word than its other form *tabair*. And, although the preverb *do* was homophonous with the preposition *do* 'to', neither of these would be considered a whole phonological word, so an Ancient Greek-style analysis of enclitics appearing after the first word in a clause is not sufficient for this data, even overlooking examples like (20).

Clitics in Old Irish show a more limited distribution than in Pashto due to the limitations the syntax puts on what may appear clause-initially. However, both languages have clitics that seem to disrupt the integrity of certain classes of verbs, and that choose their hosts based on criteria that are not straightforwardly phonological (why can Pashto clitics appear after a whole complex DP?) or syntactic (why can't Old Irish clitics appear after "if", or Pashto ones after unstressed prepositions?).

3 Syntax of the VP in Pashto and Old Irish

Although Old Irish has VSO surface word order, I propose an underlying SOV structure for both Pashto and Old Irish. I see no empirical reason to treat Old Irish as having underlyingly head-initial or head-final VPs, but earlier Celtic languages show SOV surface order, and it seems reasonable to assume that proto-Irish was a continuation of that. Using the same structure for both languages will also simplify the current analysis.

Following Adger (2006), I assume a Larsonian VP shell (Larson 1988) for compound verbs in Old Irish and propose extending it to a-initial and compound (Classes II and III) verbs in Pashto. The structure for both languages will appear as in (21):



A structure which assumes a vP -generated subject is also possible, and I think it makes no difference in my analysis either way. I also assume that the subject enclitics (for Pashto) and object enclitics (for both languages) are generated in the VP for analysis at LF by the semantics, but a structure with coindexed *pro* forms and a separate clitic projection is certainly possible, and could help explain examples of clitic doubling in Old and Middle Irish, which I do not specifically deal with here.

In order to derive the generalization that nothing but a clitic may appear between the preverb and verb in either Pashto or Old Irish (and that they seem to form one word), I require that both the subject and object move out of the VP. For Old Irish, this may not be strictly necessary, as the verb and preverb will both raise to sentence-initial positions and generate the correct word order and desired amount of “univerbification”. The exact nature of this movement for Pashto is outside the scope of the present paper, but would be necessary for a more thorough analysis of the problem.

The reason for this movement, besides generating correct word order, and the difference between Pashto class II and class III verbs is explained in the mapping from the syntax to the prosody in section 4.

4 Syntax-Prosody Mapping

The mapping of the syntactic structures into phonological phrases may be done in Optimality Theory with the use of three constraints, details of which can be found in Selkirk (1995). They are as follows:

(22) $\text{ALIGN}[\text{XP-L}]_{\Phi}$ = Every XP aligned to the left edge of a phonological phrase

(22) is a family of constraints ranked for how strongly that lexical category prefers to be aligned to the left edge of a phonological phrase. For VP, DP, and AdvP, this will be highly ranked, but for NP, AdjP and PP, it will be ranked below the WRAP constraint. This will ensure that VP, DP, and AdvP begin their own phonological phrases, but NP, AdjP, and PP are contained in the same phonological phrases as the heads they appear as complements to.

(23) WRAP-XP_{Φ} = The right and left edges of an XP correspond to the right and left edges of a phonological phrase

This ensures that heads and their complements are in a phonological phrase together and will allow for very long DPs in Pashto to be a single phonological phrase.

(24) $\text{NORECURSION}_{\Phi}$ = There may not be a phonological phrase within another one.

The constraint in (24) is very highly ranked, and will prevent the left-edge align constraints from embedding phonological phrases within other phonological phrases.

For the data I am analyzing, the same ranking of constraints will hold for both Pashto and Old Irish, but it is possible that more detailed study into the prosodic structures of both languages would yield a different analysis. The ranking I use is in (25):

(25) $\text{NORECURSION}_{\Phi} > \text{ALIGN}[\text{VP-L}]_{\Phi}, \text{ALIGN}[\text{DP-L}]_{\Phi} > \text{WRAP-XP}_{\Phi} > \text{ALIGN}[\text{NP-L}]_{\Phi}, \text{ALIGN}[\text{ADJP-L}]_{\Phi}$

The other reason for moving the subject and object out of the VP in Pashto was so that they would not be included in the same phonological phrase as the preverb or verb by the constraints above. A few examples of this ranking and the prosodic structures it generates are below.

(26) is a sentence with an a-initial verb and no clitics, just to show what the mapping will do. The object appears before the subject, so it must have moved out of the VP, and presumably both did, though it is not clear to me exactly what syntactic positions either should occupy – perhaps a TopicP for the object, and Spec,TP for the subject. For the present analysis, I will look at an extremely simplified syntax which ignores phonologically empty projections like CP and TP (though there does seem to be tense on the verb, I am assuming the verb remains in the VP).

- (26) kor fpanə axli (Kaisse 1981:202)
 house shepherd buys
 'The shepherds are buying the house.'

I assume that the syntax which is fed into the prosody appears something like in (27):

- (27) [DP_{[NP kor]]][[DP_{[NP fpanə]]][VP a_[VP xli]]}}

Although there is evidence that traces and possibly other phonologically null structure may be visible at the syntax/phonology interface, I am ignoring them for the sake of simplification of the present analysis. The mapping of (27) into phonological phrases will work as in (28) below:

(28)

[DP _{[NP kor]]][[DP_{[NP fpanə]]][VP a_[VP xli]]}}	NOR	ALIGN [VP-L] _Φ	ALIGN [DP-L] _Φ	WRAP-XP _Φ	ALIGN [NP-L] _Φ
a. ((kor)φ)((fpanə)φ)(a(xli)φ)φ	*!***				
b. ((kor fpanə)φ)(a)φ(xli)φ			*!	*	*
c. (kor)φ(fpanə)φ(a xli)φ		*!		*	**
☞ d. (kor)φ(fpanə)φ(a)φ(xli)φ				*	**

In (28a), every syntactic phrase is mapped to its own phonological phrase, at the cost of violating NoRecursion heavily. In (28b), there is no recursion, but the second DP does not align with the left edge of a phonological phrase, and it incurs a fatal violation there. Similarly, (28c) does the same thing, but violates it at the VP level rather than the DP. The optimal candidate, (28d), is the one which groups the embedded NPs into phonological phrases with the (null) DPs they are complements to, but creates a new phonological phrase at each DP and VP, so that the three word sentence in (26) maps onto four phonological phrases.

Similar prosodic structures will be derived for Old Irish, as we can see in a sentence based of the one in (15), but using a compound verb:

- (29) Do -beir in fer in claideb
 PREVERB bring.3SG the man the sword
 'The man brings the sword.'

The syntax of this will be fed into the prosody in some form like in (30):

- (30) [VP do_{[VP beir]] [DP in_{[NP fer]] [DP in [NP claideb]]}}

The mapping will insert left edges of prosodic phrases before each VP and DP, and the NPs will be included in the same phrase as the DPs, for a prosody as in 31:

(31) (do)ϕ (beir)ϕ (in fer)ϕ (in claideb)ϕ

While mapping the preverbal particles in both languages onto their own prosodic phrase (even if we were to say it is a minor prosodic phrase, in a more complicated analysis of the prosody), but there is phonological evidence in Old Irish for this split beyond just the placement of enclitics.

In Old Irish, there are vowel-final preverbs which appear before vowel-initial verbs. There is never any elision of these vowels in the manuscripts, but there is elision of the final vowel of a possessive pronoun before a vowel initial noun.

(32)a. do-icc (*d'icc)
 'comes'

(Green, 1995:47)

b. mo ort → m'ort
 1sg rank
 'My rank'

(Thurneysen 1961:277)

This supports my proposed prosodic structure with a prefix in a separate phonological phrase from the verb, but where a possessive pronoun would share a phrase with its noun complement. We can say, fairly uncontroversially, that this elision does not occur across prosodic phrase boundaries.

4.1 Pashto DP subjects

While my analysis creates a lot of very small phonological phrases, it will also allow very large phrases. The ranking of WRAP relative to the family of ALIGN constraints means that a DP or VP will have all of its complements in the same phonological phrase as it. This means that examples such as (3a), repeated below in (33) will have the entire first DP in one phonological phrase.

(33)a. [NP aya fel kaləna danga aw khaysta peyla] me nən bya wəlida (Tegey 1977:83-4)
 that 20 year tall and pretty girl I today again saw
 'I saw that twenty-year-old tall and pretty girl again today.'

b. [DP aya [AdjP fel [NP kaləna] [AdjP danga [&aw [AdjP khaysta [NP peyla]]]]] me [AdvP nən] [AdvP bya] [VP wə [VP lida]]

c. (aya fel kaləna danga aw khaysta peyla)ϕ me (nən)ϕ (bya)ϕ (wə)ϕ (lida)ϕ

The syntax in (33b) maps onto the prosody in (33c), and the clitic simply attaches to the first prosodic phrase in the clause. A possible problem in the way I have defined and ranked the constraints is that a DP which included an embedded VP or another DP would trigger the insertion of a new phonological phrase, though the small amount of evidence I have for such structures in Pashto suggests that these do not create new phrases. The example from (10), repeated in (34), has a coordinated structure with two personal names, each of which would presumably be in a DP, but the clitic attaches after the entire coordinated DP phrase. Additional data on clitic placement in sentences

of the types described here would be helpful in refining my analysis to more accurately map the syntax onto the prosody.

(34)a. [_{DP} xufal aw patang] **be ye** dər ta rawɿi (Tegey 1977:83-4)
Khosal and Patang will it you to bring
'Khosal and Patang will bring it to you.'

b. [_{DP} xufal [& aw [_{DP} patang]]]

c. (xufal aw)ϕ (patang)ϕ

The prosodic structure in (34c) is not desired, because it would predict that the enclitics would attach after *aw* 'and', instead of after *patang*, as is attested.

4.2 Pashto verb classes

The last piece of evidence I will look at in this section is what implications this prosodic structure has for the Pashto verb classes. As a review, class I is monomorphemic verbs, but includes the a-initial verbs, which can be split by an enclitic. Class II is compound verbs of a preverb+verb stem, and class III is compound verbs of a lexical item + verb stem.

Roberts (2000) expresses some concern about a phonological analysis of Pashto clitic placement involving a phonological phrase host that includes nothing more than a stressed vowel. While I hope that the comparison with Old Irish would alleviate this, it is also worth discussing the ways in which mapping to prosodic phrases is different from mapping to prosodic words, and that the one need not necessarily include the other.

If we assume, as explained in section 4 above, that mapping to prosodic phrases looks only at the syntax, then whether a constituent includes only a stressed vowel, or a fairly complex DP, that constituent will be made to form a single prosodic phrase. Prosodic words, we can say, again following Selkirk (1995), are determined by the lexicon. A constraint like (35) should work:

(35) $Lx \approx \omega$: A lexical category must form a prosodic word; a non-lexical category must not form a separate prosodic word.

With this added constraint (and likely a few others) we can further refine the prosodic structure of these languages by looking at the prosodic word level as well as phrases. In Pashto, the a-initial verbs and class II verbs will have prefixes that are not words, and so their phonological phrases will not contain phonological words. The class III verbs, however, do contain lexical words and auxiliary verbs, so these will be two separate phonological phrases and two separate phonological words. Some examples of my proposed structures are in (36).

- | | | |
|--------|---|-----------|
| (36)a. | [[skundólə] _ω] _φ
'was pinching' | Class I |
| b. | [(a) _φ (xistələ) _φ] _ω
'buying' | A-initial |
| c. | [(tel) _φ (wahó) _φ] _ω
'was pushing' | Class II |
| d. | ([wobə] _ω)([kávəl] _ω) _φ
'is watering' | Class III |

The differences between the classes of verbs is shown in their prosodic structures. Class I consists of a phonological phrase which coincides with a phonological word. The a-initial and Class II verbs consist of two phonological phrases within a single phonological word, and Class III verbs are two phonological phrases which coincide with two phonological words. I believe it is extremely typologically strange and potentially theoretically problematic to have a phonological word be larger than a prosodic phrase, but this would be a matter for future research, or simply the redefining or creating of terms.

4.3 Summary

The syntax-prosody mapping occurs in such a way that morphologically complex verbs, because they contain more than one VP, are mapped into two separate phonological phrases, despite containing only one phonological/lexical word. Complex DPs, on the other hand, map into a single phonological phrase and act as a constituent in the phonology as well as the syntax. These facts will form an important distinction for the placement of clitics in the next section.

5 Clitic placement with Optimality Theory

With a basic idea of how the syntax looks in Pashto and Old Irish verbs and how the syntax/prosody mapping works, we can now look at an Optimality Theory account of how the phonology places the clitics. In addition to providing rankings for Pashto and Old Irish, I will also sketch the beginnings of what a typology could look like by accounting for some basic facts about Serbo-Croatian and Ancient Greek clitics as well.

A small number of constraints should be sufficient to place second-position clitics. First is a constraint that they be as far left in the clause as possible. Roberts (1997:371) suggests that for Pashto, this is a family of ranked constraints for each specific clitic in order to derive the ordering from (7) that is seen when more than one clitic appears in a clause. This approach seems to work well, and is more detailed than this analysis needs at present. In Old Irish, there are only object clitics, and no double-object constructions, so clitic ordering is irrelevant. The *EDGEMOST* constraint for this analysis appears in (37).

(37) EDGEMOST[CL-L] = Clitics must be leftmost in their clause

Because second position clitics never appear leftmost in their clause, the constraint in (37) is outranked by a constraint that clitics have an appropriate host:

(38) LEAN[CL-L] = An enclitic must lean left onto a stressed host

Roberts (1997:390) uses a constraint on phonological phrase prominence to accomplish the same, with the goal of ruling out the unstressed prepositional phrases as host candidates. The phrasing of his constraint however, might pose a problem for the Old Irish data. It suggests that the clitics might (in some language, if not in Old Irish), move rightward to lean onto a more prominent host. That is, that Old Irish clitics might prefer the heavily stressed verb stem as a host over the more lightly stressed preverbal particle or negation. For (38), any amount of stress will do, and clitics will not try to attach to the most prominent stress in an utterance.

(39) ALIGN[CL-L, Φ -R] = Align the left edge of a clitic to the right edge of a phonological phrase

The constraint in (39) seems suited to derive the placement of Pashto clitics, but for Old Irish, it is probably more accurate to align the right edge of a clitic to the right edge of a phonological phrase for a “nesting” effect. Because there is some degree of phonological assimilation between the host and the clitic (depending on one's analysis of the allomorphy, which I have glossed over in this paper) it makes sense to position the clitic in a way that will not have a phrase boundary blocking these phonological processes.

(40) ALIGN[CL-L, ω -R] = Align the left edge of a clitic to the right edge of a phonological word

The constraint in (40) works in the same way as the one in (39), but has a different target size for the host of a clitic. This will allow us to derive Ancient Greek and the optionality of hosts in Serbo-Croatian.

(41) INTEGRITY ϕ/ω = Do not insert phonological material into a prosodic phrase/word

Roberts (1997) uses this as a constraint that is violated by Serbo-Croatian at the phrase level, and by Pashto at the word level. I think that having the align constraints select the appropriate host, one can get away without using this pair of constraints.

5.1 Pashto and Old Irish

For these languages, the constraint ranking is as follows:

(42) LEAN[CL-L] > ALIGN[CL-L, Φ -R] > EDGEMOST[CL-L] > ALIGN[CL-L, ω -R]

This means that a clitic selecting a host will care most about its host having stress, then that

the host is the right edge of a prosodic phrase, then that it be as far left as possible, and it will not be very concerned at all about its host being a prosodic word. This can be seen in the tableau in (43) with the example from (14c), repeated below:

- (14c) [PP ra ta] [PP te] [AdvP ra] [kawə́] **de** (Tegey 1977:119)
 me for from-it here pick you
 'You were picking it for me from it (and bringing it) here.'

(43)

(ra ta)φ (te)φ (ra)φ ([kawə́]φ) + de	LEAN [CL-L]	ALIGN [CL-L, ϕ-R]	EDGEMOST [CL-L]	ALIGN [CL-L, ω-R]
a. (ra ta)φ de (te)φ (ra)φ [([kawə́]φ)ω]	*!		*	*
b. (ra de ta)φ (te)φ (ra)φ [([kawə́]φ)ω]	*!	*		*
c. (ra ta)φ (te)φ (ra)φ de [([kawə́]φ)ω]	*!		***	*
☞ d. (ra ta)φ (te)φ (ra)φ [([kawə́]φ)ω] de			****	

While every candidate but the winner failed right away for violating LEAN, we can also look at the other violations they show. (43b) has a clitic that is not aligned to the right edge of a prosodic phrase, and incurs a violation mark. I can not think of an example where it would be helpful to penalize this twice by having an INTEGRITY constraint. EDGEMOST was violated by every candidate that had a potential landing site (as defined by ALIGNφ) to its left. Another possibility for counting violations against it would be to mark against every word or phrase, but this won't have any effect on the results. The winning candidate violated EDGEMOST more than the others, but is still optimal for being the best compromise between EDGEMOST and lean. (4d) also happens not to violate ALIGNω, but this constraint will likely never decide between two candidates in Pashto.

The results in Old Irish should look similar. Here, I use the sentence from (20).

- (20) ma immi -**m** 'thabarthar, (Milan Glosses, 41^c2)
 if PREVERB -1SG surround.3SG.present.deponent
 'If I am surrounded,'

(44)

(ma)φ [(immi)φ ('thabarthar)φ)ω + - m	LEAN [CL-L]	ALIGN [CL-L, ϕ-R]	EDGEMOST [CL-L]	ALIGN [CL-L, ω-R]
a. (ma)φ - m [(immi)φ ('thabarthar)φ)ω	*!		*	*
☞ b. (ma)φ [(immi)φ - m ('thabarthar)φ)ω			**	*
c. (ma)φ [(immi)φ ('thabarthar)φ)ω - m		*!	***	

In (44), the winning candidate is the one that has a host with (even a secondary) stress. (44c)

does not gain anything by having a “better” stress, or by attaching at a word boundary rather than a phrase boundary.

5.2 Ancient Greek

The difference between Ancient Greek and Pashto/Old Irish is that the clitics attach to a host that is a phonological word. It is easy enough to re-rank the constraints to derive this pattern. A possible ranking for Greek is in (45):

(45) LEAN[CL-L] > ALIGN[CL-L, ω-R] > EDGEMOST[CL-L] > ALIGN[CL-L, Φ-R]

Now, the clitic will be looking for a word as a host to attach to, rather than a phrase. As I have only the example from the introduction in (1) and it is not particularly exciting, I will assume it is possible to understand the above description and generalization without a tableau.

5.3 Serbo-Croatian

Here, there is more to say. The optionality between word hosts and phrase hosts means that the two align constraints must be equally ranked in Serbo-Croatian, allowing speakers to choose between candidates that are otherwise equal. For this, our ranking will be as in (46).

(46) LEAN[CL-L] > ALIGN[CL-L, ω-R], ALIGN[CL-L, Φ-R] > EDGEMOST[CL-L]

This is illustrated in the tableau in (46) with the example from (2). I am guessing at the prosodic structure of Serbo-Croatian, but the crucial part is that the name is two prosodic words contained in one prosodic phrase, which should not be controversial.

(2)a. [DP Lav Tolstoi] **je** veliki ruski pisac (Halpern, 1995, via Roberts 1997:380)
 Leo Tolstoy is great Russian writer

b. [Lav]_ω **je** Tolstoi veliki ruski pisac
 Leo is Tolstoy great Russian writer
 'Leo Tolstoy is a great Russian writer.'

(46)

[[Lav] _ω [Tolstoi] _ω] _φ ([veliki] _ω [ruski] _ω [pisac] _ω) _φ + je	LEAN [CL-L]	ALIGN [CL-L, ω-R]	ALIGN [CL-L, Φ-R]	EDGEMOST [CL-L]
a. ([Lav] _ω [Tolstoi] _ω) _φ ([veliki] _ω [ruski] _ω [pisac] _ω) _φ je		*		*****!
☞ b. ([Lav] _ω je [Tolstoi] _ω) _φ ([veliki] _ω [ruski] _ω [pisac] _ω) _φ			*	*
☞ c. ([Lav] _ω [Tolstoi] _ω) _φ je ([veliki] _ω [ruski] _ω [pisac] _ω) _φ		*		*
d. je ([Lav] _ω [Tolstoi] _ω) _φ ([veliki] _ω [ruski] _ω [pisac] _ω) _φ	*!			

In (46), the candidates showing the two attested positions for the clitic come out tied. In this case, we see it is crucial how one counts up violations against EDGEMOST. If one counts spaces from the left edge as determined by how many words or phrases (depending on which was selected as a host), then (46b) and (46c) each have one violation, and are perfectly tied.

6 Conclusion

In this paper, I have introduced the complicated and confusing facts of clitic placement in Pashto and Old Irish, and have proposed a phonological analysis that I believe explains the data. Clitics in these languages search for a prosodic phrase as a host, and attach to the first stressed one in their clause. The complications arise from how the syntax/prosody interface determines what the domain of a phonological phrase is, and how this differs from a phonological or lexical word. The fact that there are two, not particularly closely related languages that seem to behave in a very similar way in this regard should ease worries of its typological rarity. There are still some details to be worked out regarding the overlapping of phonological phrases and words and DPs which contain other DPs or VPs, but the analysis also lends itself to languages showing different patterns of clitic placement. Future research should look more closely at the theoretical issues raised here, specifically in what an analysis of this type says about the Strict Layer Hypothesis (Selkirk 1984:26), the minimum size of a phonological phrase, and the integrity of lexical words. Possible avenues for bringing this analysis in line with previous research include the following:

- I. Assume that the Strict Layer Hypothesis holds and is theoretically desirable. Rephrase the analysis above in a minimal way such that the lexical/semantic words *dobeir* and *axistəl*, for example, still consist of two phonological phrases, (allowing the crucial parts of my analysis to hold) but only *-beir* and *-xistəl* are phonological words, plus the augments *do-* and *a-*. This assumes that Pashto and Old Irish have no minimum size requirement for phonological phrases (similar perhaps to French, where phonological phrases do not have to branch (Inkelas & Zec, 1995:545). This also assumes that the lexicon has different criteria for word-hood than is strictly phonological.
- II. Assume that the Strict Layer Hypothesis is either not active in all languages, or is a violable constraint. This could provide evidence for a direct access theory instead of a Prosodic Hierarchy Theory. Inkelas and Zec point out that it is difficult to obtain empirical data for these competing theoretical predictions due to the relatively “small number of (described) postlexical phonological rules operating in subutterance domains in a single language.” (1995:548)
- III. Keep the assumptions and mapping as suggested in the current paper, and look for more examples of languages where a similar approach is helpful. Inkelas and Zec mention that “preliminary work on highly agglutinating languages has suggested that words, like sentences, may also be parsed into a phonological structure distinct from their morphological structure.” (1995:548) This could include lexical words consisting of more than one phonological phrase.

References

- Adger, D. 2000. First Position and the Syntax/Prosody Interface: Old Irish Preverbs, in *WCCFL 19 Proceedings*, ed. Billerey and Lillehaugen, pp. 1-14. Cascadilla Press, Somerville, MA.
- Adger, D. 2006. Post-Syntactic Movement and the Old Irish Verb. *Natural Language and Linguistic Theory* 24:604-654.
- Babrakzai, F. 1999. Topics in Pashto syntax. PhD dissertation, University of Hawai'i, Manoa.
- Carnie, A., Harley, H. & Pyatt, E. 2000. VSO Order as Raising Out of IP? Some Evidence from Old Irish. In *The Syntax of Verb Initial Languages*, eds. Carnie, A. and E. Guilfoyle.
- Dost, A. 2007. Linearization, square pegs, and round holes. PhD dissertation, University of California, Santa Cruz.
- Eska, J. 2009. Where have all the object pronouns gone? The growth of object agreement in earlier Celtic. *Zeitschrift für Celtische Philologie* 57. pp. 25-47.
- Green, A. 1995. *Old Irish Verbs and Vocabulary*. Cascadilla Press, Somerville, MA.
- Halpern, A. 1995. *On the placement and morphology of clitics*. CSLI Publications, Stanford.
- Inkelas, S. & D. Zec. 1995. Syntax-Phonology Interface. In *The Handbook of Phonological Theory*, J. Goldsmith, ed. London: Blackwell . pp. 535-549.
- Kaisse, E. 1981. Separating phonology from syntax: A reanalysis of Pashto cliticization. *Journal of Linguistics* 17:197-208.
- Kern, G. 2010. On Secondary Stress in Old Irish. MA thesis, University of Wisconsin-Madison.
- Larson, R. 1988. On the Double Object Construction. *Linguistic Inquiry* 19:335-391.
- Nespor, M. & Irene Vogel. 1986. *Prosodic Phonology*. Foris Publications, Dordrecht.
- Quin, E.G. 1975. *Old-Irish Workbook*. Royal Irish Academy, Dublin.
- Roberts, T. 1997. The optimal second position in Pashto. In *Phonology in progress—progress in phonology*, eds. G.E. Booij and Jeroen Maarten van de Weijer, volume 3 of *HIL Phonology Papers*, 367-401. Holland Institute of Generative Linguistics, Amsterdam.
- Roberts, T. 2000. Clitics and Agreement. PhD dissertation, Massachusetts Institute of Technology.
- Selkirk, E. 1984. *Phonology and Syntax*. MIT Press, Cambridge, MA.
- Selkirk, E. 1995. Sentence prosody: intonation, stress and phrasing. In *The Handbook of Phonological Theory*, ed. John Goldsmith. London: Blackwell. pp. 550-569.
- Strachan, J. 1949. *Old Irish Paradigms and selections from the Old Irish glosses*. Royal Irish Academy, Dublin.
- Tegey, H. 1977. The grammar of clitics: Evidence from Pashto and other languages. PhD dissertation, University of Illinois, Urbana-Champaign.
- Tegey, H. & B. Robson. 1996. *Pashto reference grammar*. Center for Applied Linguistics, Washington, D.C.
- Thurneysen, R. 1961. *A Grammar of Old Irish*. Dublin Institute for Advanced Studies, Dublin.
- Wackernagel, J. 1892. Über ein Gesetz der indogermanischen Wortstellung. *Indogermanische Forschungen I*, 333-436.